Gesture Recognition: Translating Hand Symbols into Meaningful Names

Sujitha Chintalapudi[1], Karri Vasundhara[2], Sarayu Potnuru[3],
Bhagyasri Raja Rajeswari Devi Sandanala[4], Adiraju Raj Narayan Kaushik[5] and
B.C.S.N. Murthy Nukala[6]
{sscp929@gmail.com[1], karrivasundhara10@gmail.com[2],swetapotnuru83@gmail.com[3]
bhagyasrisadanala@gmail.com[4], kaushik.adiraju@aditya.ac.in[5], nmurthy@aditya.ac.in[6]}

Department of BCA Data Science, Aditya Degree & PG College, Kakinada (Autonomous),
Andhra Pradesh, India[1]
Department of BSc Data Science, Aditya Degree College, Tuni, Andhra Pradesh, India[2]
Department of B.Sc Data Science, Aditya Degree College, Gajuwaka, Andhra Pradesh, India[3]
Department of B.Sc Data Science, Aditya Degree & PG College for Women's Kakinada (Autonomous),
Andhra Pradesh, India[4]
Assistant Professor, Department of B.Sc Data Science, Aditya Degree & PG College, Andhra Pradesh,
India[5]
Associate Professor, Department of B.Sc Computer Science, Aditya Degree & PG College, Kakinada,
Andhra Pradesh, India[6]

Abstract. This paper introduces a novel approach to gesture recognition tailored to empower individuals with hearing impairments. Leveraging a hybrid CNN+LSTM architecture, our model learns directly from user-provided gestures, bypassing the need for pre-existing datasets and enabling real-time adaptation. Through comprehensive experiments, we demonstrate the efficacy of our approach in accurately translating hand symbols into meaningful names. Our findings underscore the potential of personalized gesture recognition systems in fostering accessible and intuitive communication for the hearing impaired.

Keywords: Gesture recognition, CNN, LSTM, Hybrid model, Cross-communication, Hearing impaired, Real-time adaptation, Hand symbols.

1 Introduction

In times we live in where Technological Development is sideward in brakes on what it comes to accessibility and communication the importance of addressing a variety of users to the user diversity should be in the spotlight. Solving the communication for the hearing impaired, this project is set to change cross-communication through GESTURE RECOGNITION: An innovative solution far beyond the boundaries of languages. The task, which should be addressed with high level considerations to the adaptation in real time and to the personalized interaction, exploits cutting-edge DEEP LEARNING techniques, and it is mainly based on the CNNs and LSTM networks. The combination of CNNs and LSTMs forms a HYBRID MODEL-or simply HYBRID, that decodes and translates hand symbols in a fashion that appeals to the distinctive requirements of the hearing impaired. Using flash 3d in combination with arrows and state of the art frameworks like TENSORFLOW and MEDIAPIPE, the project captures the spirit of present machine learning and computer vision model, allowing to recognition of robust and efficient gesture in real environment. The workflow of the project

ICITSM-Part II 2025, April 28-29, Tiruchengode, India
Copyright © 2025 EAI
DOI 10.4108/eai.28-4-2025.2358117

proceeds methodically starting from the robust capture of gesture data through a carefully constructed pipeline. Leveraging the powerful features and functionalities provided by MEDIAPIPE's modular infrastructure, hand gestures are recognized and analyzed, and as a result, a dataset with rich manual annotations (i.e., including gesture labels) is created. The later steps of the project highlight the hybrid model design (The complex architecture including the CONVOLUTIONAL and RECURRENT neural network layers). This ARCHITECTURE enables easy feature extraction and temporal modeling to allow the model to capture the complex details present in the gestures. Moreover, the importance of the project extends beyond the reached technical feats, it also encompasses a wider vision of inclusivity and outreach. Through avoiding dependence on legacy datasets and adopting a persona-centric inter- if the project's commitment to personalized interaction and real- world utility may be highlighted action design, moreover. The inclusion of functionalities such as real-time gesture-based recognition and no-profile device-to-device communication sys- tem represents a paradigm shift in ASSISTIVE TECHNOLOGY This bears a promise for a future where hearing impaired people can interact with their environment more intuitively.

2 Literature Review

Miah et al. [1] proposed an innovative dynamic hand skeleton based hand gesture recognition methodology. They introduced a Multi-branch attention-based graph + general deep learning. The proposed method obtained high accuracies on the benchmark datasets compared to the prior art methods.

Hax et al. [2] introduced a hybrid architecture for recognizing dynamic hand gestures in uncontrolled environments. Their RNN+CNN model achieved a mean accuracy of 83.66 %.

Kabir et al. [3], CSI-DeepNet, a light-weight deep learning-driven system for hand gestures was introduced. They reached an 96.31% accuracy using CSI and low-power system- on-chip ESP-32. This method achieves the state-of-the-art performance more accurate and more computationally efficient than traditional CNN-based methods.

R. Suguna et al. [4] presented were a detection and recognition method for human hand gestures and the possible uses cases in the field of automated vehicle activity. The method consists of taking a classification deep learning model, CNN-based, for multiple steps in the flow of the method. Firstly, we segmented hand regions of interest based on mask images, then we obtained finger segmentation and normalized the segmented finger images. The segmentation uses adaptive histogram equalization for contrast improvement of the image. Also, the connected component analysis is employed to separate the finger tips from the hand images.

Ghulam Muhammed et al. [5] introduced the importance of hand gesture recognition, stating that it has a wide range of application fields, from video game to telesurgery, and it works importantly for the translation of sign language. But sign language (with hand shapes and gestures) depends on formations, orientations, and positional relations for structured expression.

Reena Tripathi et al. [6] fine-tuned experiments on the Cambridge Hand Gesture Dataset (CHG [9]) using CLIP-LSTM and obtained the accuracy of 97.0%. The CHG dataset A dataset of 900 image sequences of 9 categories of hand gestures and each gesture category is performed with 3 primitive hand shapes and 3 motions. Using the CLIP model for feature extraction and LSTM for classification the authors experimentally demonstrated the ability of their method to classify the hand gestures correctly in this difficult dataset.

Maricel L. Amit et al. [7] presented a System with LSTM and MLP to identify hand gesture using MediaPipe Holistic Model. The data set was 7000 samples of a single signer with 30 video sequences and 30 frame length in the seven sign classes employed via MediaPipe Holistic. The number of samples per class was 950 for training, with the remaining 50 samples for testing. In real time hand gesture recognition, over the course of 1000 epochs, 100% accuracy was achieved by the LSTM architecture.

Peijun Bao et al. [8] proposed the study on visual hand-gesture recognition with a dataset comprising seven classes of gestures, totaling 500,000 hand gesture samples. They utilized a deep convolutional neural network, along with AlexNet and VGG19 models, achieving 97.1% accuracy on simple backgrounds and 85.3% on complex backgrounds with their proposed network. To address overfitting, they adopted small convolutional filter sizes of 3x3 pixels and applied a dropout learning strategy to prevent co-adaptation of network units. Additionally, an early stopping criterion for training was implemented to complement measures aimed at avoiding overfitting.

Hung-Yuan et al. [9] proposed for instant hand gesture recognition for home appliance control or human-computer interaction using a webcam. They employed skin color detection, morphology, and background subtraction to isolate the hand region, subsequently applying kernelized correlation filters (KCF) for tracking. Resized hand images were fed into modified AlexNet and VGGNet-based deep convolutional neural networks (CNNs) for gesture identification. With a training dataset achieving a recognition rate of 99.90%, and a test dataset recognition rate of 95.61%, the study demonstrated practical feasibility. Data preprocessing techniques and a dataset of 4800 training images, with varied backgrounds and angles, contributed to robust model training and verification.

Jayaprakash et al. [10] proposed two new approaches for hand gesture recognition in sign language. They introduce a hybrid feature descriptor that combines SURF and Hu Moment Invariant methods to create a feature set with both high recognition rates and low time complexity. Additionally, they enhance recognition performance and resilience to viewpoint variations by introducing derived features from the combined feature set. Their methods utilize K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) for hybrid classification of single signed letters. Furthermore, they propose finger- spelled word recognition using Hidden Markov Model (HMM) for a lexicon-based approach. Experimental results demonstrate the effectiveness of the proposed approaches, showing improved real-time efficiency and robustness compared to other popular techniques.

3 Data Description

The dataset employed in this project is meticulously curated to facilitate robust training and evaluation of the gesture recognition model. Comprising hand gestures performed by individuals with varying gestures, the dataset encapsulates a diverse range of hand symbols, each annotated with corresponding labels denoting their semantic meaning.

Each gesture sequence is recorded using a standard webcam setup, capturing 30 frames per video with a suitable resolution. The dataset encompasses 30 sequences, with each sequence corresponding to a unique gesture performed by the participant. Notably, the dataset encompasses gestures relevant to cross-communication, with emphasis on expressions commonly utilized by individuals with hearing impairments. Fig 1 Shows the Data: Hand Gestures [11].

Fig. 1. Data: Hand Gestures [11].

At each frame of sequences, hand pose and motion is raycasted using keypoint information obtained from the Mediapipe framework. These keypoints coordinates are also included as input to the gestures recognition model (see the left and right-hand landmarks).

Moreover, the dataset is organized into separate action classes covering a range of hand gestures from simple expressions (e.g." Hello") to subtler symbols such as" Love" and" Namaste". This taxonomization provides the training and testing data for the generative model of the gesture recognition model to learn the subtle differences in the hand movements, which corresponds to semantically different meanings.

Conclusion The dataset is a fundamental source for the development of the gesture recognition field, offering to researchers and practitioners a rich corpus of hand gestures with corresponding labeled annotations. Its diversity, granularity, and natural relevance to practical communication situations make it an invaluable resource for promoting innovation in the field of assistive technology and human-computer interaction.

4 Methodology

This study's methodology includes a methodical approach intended to achieve the goals of creating a strong gesture recognition system specifically suited to assisting people with hearing impairments in cross-communication. Data collection, model construction, training, and evaluation are the many stages of the process. In order to train the gesture recognition model, the project's first phase is gathering gesture data. Mediapipe, a complete framework for real-time hand gesture detection and landmark extraction, is used in this procedure. Through the integration of a webcam, hand gestures are captured in diverse environmental conditions, ensuring the diversity and richness of the dataset. Each gesture is meticulously annotated with corresponding labels, facilitating supervised learning and model training. Fig 2 Shows the Hybrid Model architecture [12].

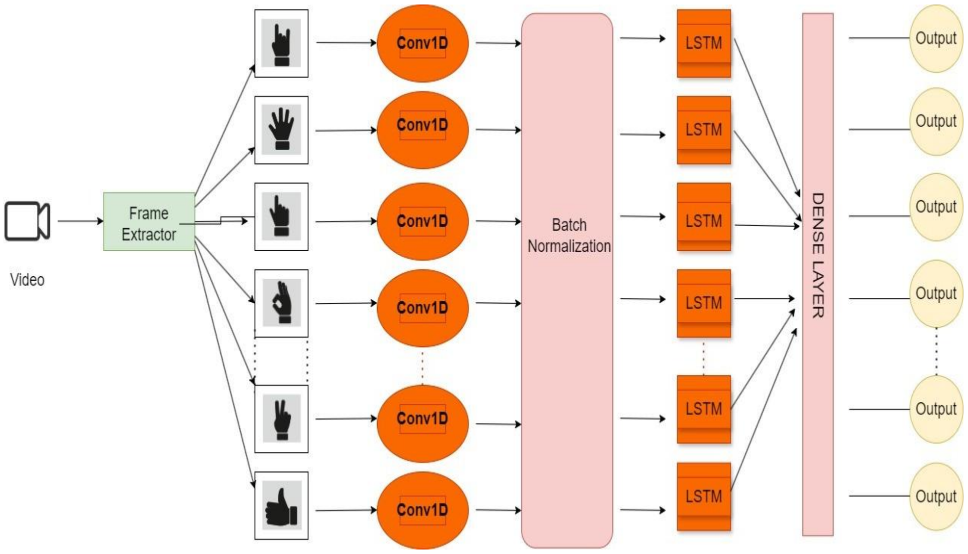
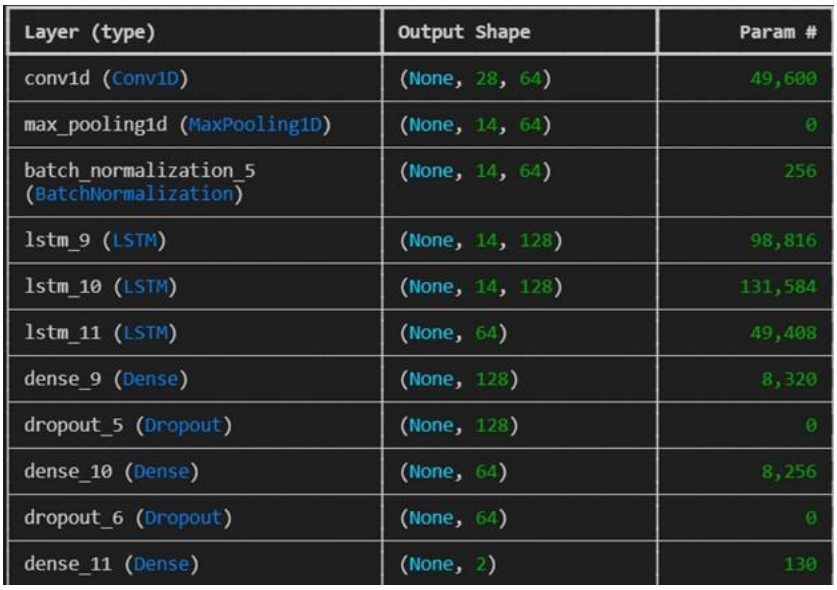


Fig. 2. Hybrid Model architecture [12].

Central to the methodology is the development of a hybrid gesture recognition model, leveraging the symbiotic relation- ship between CNNs and LSTM networks. CNN layers facilitate feature extraction, capturing spatial information from the input images, while LSTM layers enable sequence modeling, capturing temporal dynamics and context.

This hybrid model architecture combines CNN layers with LSTM layers, followed by fully connected layers for classification tasks. Layers and Configuration Shown in Fig 3.



Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 28, 64)	49,600
max_pooling1d (MaxPooling1D)	(None, 14, 64)	0
batch_normalization_5 (BatchNormalization)	(None, 14, 64)	256
lstm_9 (LSTM)	(None, 14, 128)	98,816
lstm_10 (LSTM)	(None, 14, 128)	131,584
lstm_11 (LSTM)	(None, 64)	49,408
dense_9 (Dense)	(None, 128)	8,320
dropout_5 (Dropout)	(None, 128)	0
dense_10 (Dense)	(None, 64)	8,256
dropout_6 (Dropout)	(None, 64)	0
dense_11 (Dense)	(None, 2)	130

Fig. 3. Layers and Configuration.

4.1 Convolutional Layer (Conv1D)

The first layer is a 1D convolutional layer with 64 filters and a kernel size of 3. ReLU activation is applied to introduce non-linearity.

4.2 MaxPooling1D Layer

Max pooling with a pool size of 2 is performed to reduce the dimensionality of the feature maps.

4.3 Batch Normalization Layer

Batch normalization normalizes the activations of the previous layer, improving training stability and convergence speed.

4.4 LSTM Layers

Three LSTM layers are stacked sequentially. The first two layers have 128 units each, and the third layer has 64 units.

4.5 Fully Connected Layers (Dense)

Two connected layers are stacked onto the LSTM layers. There are 128 neurons in the first Dense layer with ReLU activation. The model is also regularized by 0.5 dropout. The second Dense layer contains 64 neurons and uses an ReLU activation, followed by a dropout layer.

4.6 Output Layer

The output layer is just a Dense layer with a softmax activation (for multi-class classification).

During training, a mixed gesture recognition model is trained using the labeled dataset. Training is an iterative process to adjust parameters that minimize the loss function and improve prediction. Training is based on the most advanced deep learning frameworks such as TensorFlow in order to accelerate model convergence and to guarantee scalability. And to prevent overfitting and to improve generalization, dropout regularization [12] and batch normalization [13] are used as well.

5 Results and Discussion

The ready model, hybrid LSTM CNN model, is tested on Seven actions dataset: Rock, Ok, Like, Dislike, Victory, Namaste and Hug dataset. The model attained a high accuracy of 98.15% during evaluation, and the corresponding loss was 0.0771. Fig 4 Shows the Training Loss.

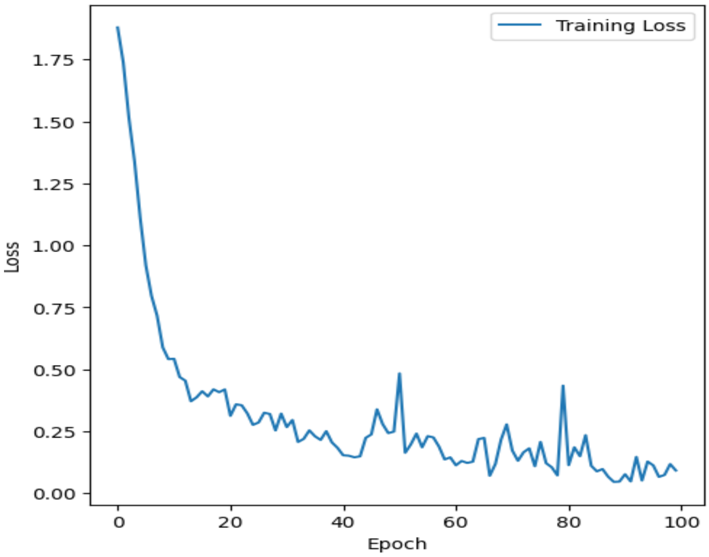


Fig. 4. Training Loss.

This high level of accuracy indicates that the model successfully learned and generalized features of the hand gesture data, and can also classify completely new samples accurately. The experimental results show that the hybrid LSTM-CNN method is effective in extracting spatial and temporal dependencies in data, complementing advantages of LSTM and CNN layers in such a manner that it can classify unseen samples with a high accuracy. The experimental results also show the ability of the hybrid LSTM-CNN model to learn both spatial and temporal dependencies in the data by exploiting the advantage of the both LSTM and CNN layers. Training Accuracy Shown in Fig 5 and Table 1 Shows the Performance Metrics of Hybrid LSTM-CNN Model.

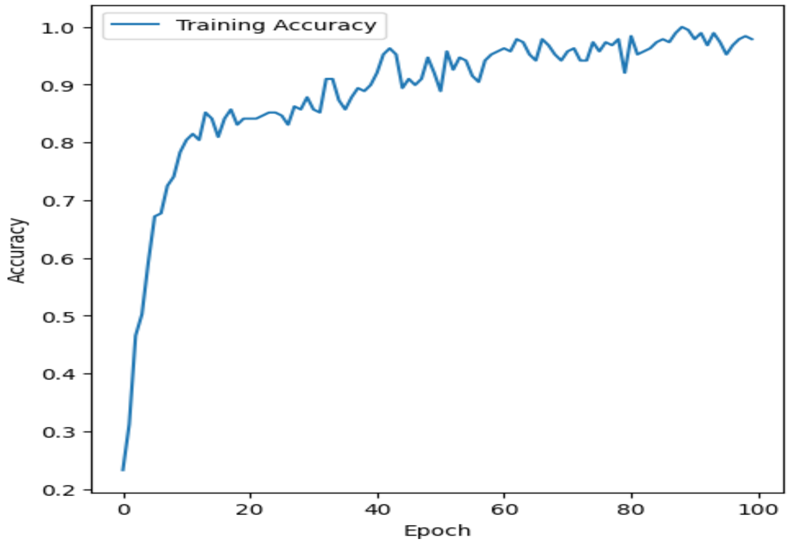


Fig. 5. Training Accuracy.

Table 1. Performance Metrics of Hybrid LSTM-CNN Model.

Metric	Value
Accuracy	0.9905
F1 Score	0.9905
Precision	0.9908
Recall	0.9905

6 Conclusions

This suggests that the learned model general is able to learn from the data of hand gesture data and generalize well to unseen cases. The experimental results demonstrate the hybrid LSTM-CNN method efficiently considers the spatial and temporal dependences by utilizing the strength of LSTM layers and CNN layers.

References

[1] A. S. M. Miah, M. A. M. Hasan and J. Shin,” Dynamic Hand Gesture Recognition Using Multi-Branch Attention Based Graph and General Deep Learning Model,” in IEEE Access, vol. 11, pp. 4703-4716, 2023, doi: 10.1109/ACCESS.2023.3235368,

[2] 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.

[3] M. H. Kabir, M. A. Hasan and W. Shin,” CSI-DeepNet: A Lightweight Deep Convolutional Neural Network Based Hand Gesture Recognition System Using Wi-Fi CSI Signal,” in IEEE Access, vol. 10, pp. 114787- 114801, 2022, doi: 10.1109/ACCESS.2022.3217910.

[4] Neethu, P. S., R. Suguna, and Divya Sathish.” An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks.” Soft Computing 24.20 (2020): 15239- 15248.

[5] M. Al-Hammadi et al.,” Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation,” in IEEE Access, vol. 8, pp. 192527-192542, 2020, doi: 10.1109/ACCESS.2020.3032140.

[6] R. Tripathi and B. Verma,” CLIP-LSTM: Fused Model for Dynamic Hand Gesture Recognition,” 2023 IEEE 20th India Council International Conference (INDICON), Hyderabad, India, 2023, pp. 926- 931, doi: 10.1109/INDICON59947.2023.10440820.

[7] M. L. Amit, A. C. Fajardo and R. P. Medina,” Recognition of Real-Time Hand Gestures using Mediapipe Holistic Model and LSTM with MLP Architecture,” 2022 IEEE 10th Conference on Systems, Process Control (ICSPC), Malacca, Malaysia, 2022, pp. 292-295, doi: 10.1109/ICSPC55597.2022.10001800

[8] P. Bao, A. I. Maqueda, C. R. del-Blanco and N. Garc´ıa,” Tiny hand ges- ture recognition without localization via a deep convolutional network,” in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 251- 257, August 2017, doi: 10.1109/TCE.2017.014971.

[9] H. -Y. Chung, Y. -L. Chung and W. -F. Tsai,” An Efficient Hand Gesture Recognition System Based on Deep CNN,” 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, VIC, Aus- tralia, 2019, pp. 853-858, doi: 10.1109/ICIT.2019.8755038.

[10] Rekha, J., J. Bhattacharya, and S. Majumder. ” Hand gesture recognition for sign language: A new hybrid approach.” In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), p. 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2011,

[11] https://www.istockphoto.com/vector/hand-gestures-line-icons-editable- stroke-pixel-perfect-for-mobile-and-web-contains-gm1192922635- 339118029,

[12] https://www.researchgate.net/figure/Architecture-of-the-Hybrid-1D- CNN-LSTM-model-for-human-activity-recognition*fig*4$_3$43341551