Recognition of Gujarati Sign Language Alphabets Using LSTM Deep Learning Approach

Nasrin Aasofwala[1], Abisek Panigrahi[2] , Shanti Verma[3], Rinkal Sardhara[4], Kalyani Patel[5]

[1,3,4]Computer Applications and Information Technology Department, LJ University, Ahmedabad Gujarat, India.
[2]SS & C Intralinks, Boston, USA.
[5]Gujarat University, Ahmedabad, Gujarat, India.
Email: nasrin.mscit16@gmail.com[1],abpani1994@gmail.com[2],verma.shanti@gmail.com[3] rinkal.sardhara@ljku.edu.in[4], patelkalyani05@gmail.com[5]

Received: 13.07.2024 Revised: 10.08.2024 Accepted: 13.09.2024

ABSTRACT
Worldwide sign languages differ, and there isn't one universal sign language. Each country and state may have its sign language or set of related sign languages. Some of the previous research studies recognized the signs, but they required instruments like gloves, sensors, and kinetics, or many other hardware instruments that are not easily accessible for everyone. In this modern era, cameras are widely used or easily accessible to everyone. Recognition of Gujarati alphabet sign language with a camera presents a cost-effective technique to detect Gujarati alphabet signs. This research contains data acquisition, image preprocessing, feature extraction, and sign recognition. Data are collected from images taken from different people at different angles with signs. Augmentation of Data technique is also used to increase the sample size of the dataset. The model which is proposed is used as a long short-term network to translate sign language with around 98% accuracy. This study contributes to the development of an effective human-machine solution for the deaf society.

Keywords: Deep Learning, Gujarati Sign Language, Long-Short Term Memory Neural Network (LSTM), Data Augmentation

1. INTRODUCTION

In India, there are several regional sign languages, including Indian Sign Language (ISL), which is used in many parts of the country, and regional variations like GSL (Gujarati Sign Language) in Gujarat [18]. The population of Gujarat State exceeds 6 crores, surpassing that of 193 out of 216 countries worldwide, constituting 90% of nations. Notably, 14% of the residents in Gujarat experience deaf-dumb disabilities [21]. Like other sign languages, GSL has its grammar rules, syntax, and vocabulary. GSL has its own set of alphabets that has a different sign for each alphabet [19]. Recognition of sign language has been a major research idea for numerous years. American Sign Language (ASL) has been explored and developed more for sign language recognition.[21]. The ASL dataset is available on Kaggle. An automated translator system designed to interpret sign language and convert it into a comprehensible form represents a powerful solution for reducing communication barriers within society. This innovative tool has the potential to help education institutes learn the alphabet in a better way so students can easily understand.

2. RELATED WORK

There are several national and international sign languages available. Around 400 articles were available in the surroundings of the language recognition system for various Indian regional and international languages. Ibrahim represents in his paper the Arabic sign language recognition system (ArSLRS) that translates Arabi words signs to text. ArSLRC collected 450 videos for different 30 words and provided a recognition rate of around 97% [1]. The Bhutanese sign language (BSL) dataset contains 20000 images of 10 BSL digits. The available BSL system used a CNN model to convert images into sign language [2]. Many image-capturing devices are used for taking images or video, like the Kinect, camera, gloves, leap motion controller, histogram equalization (HE), CLAHE, logarithmic transformation, etc. Also used is image restoration and image enhancement are used for image preprocessing [3]. This research designs an ASL learning through a game application and develops an SLR (sign language recognition system) using a leap motion controller. It uses the LSTM algorithm to recognize ASL alphabets [4]. Recognition of Sign Language for Hindi varnamala using the CNN model trains a set of 700 images per letter and a validation set of 100 images for each letter. Keras and TensorFlow are used by the model to recognize the sign [5]. Computerized recognition of sign language model that recognizes different sign gestures. Hand movement, facial expression, and body postures are added in gestures but mostly depend on hand movement. Anyone can understand sign language [6]. Halder represents in his research paper that Media pipe is used to detect complex hand gestures easily. It divides the work into three stages: a) image preprocessing, and b) media pipe is for hand landmarks. b) Data cleaning and normalization, and c) Prediction using the ML algorithm [7]. The LSTM-RNN model can be used to prepare the model using the database of videos. It converts signs into text to teach little kids about computers using sign language. [9]. Grover represents in his research paper that developing a recognition system of sign language for Hindi vowels of 6 alphabets. It uses an LSTM-CNN network or a similar architecture that uses the recognition of sign language. [10]. Wadhawan represents in his research paper the Systematic Literature Survey for SLRs. Around 400 articles were on the surroundings of the same topic. This paper also provided different sign languages like the sign language of America (ASL), the British sign language of Britain, the sign language of Arabs, etc. [11]. A research paper proposes a reciprocal sign language converter system to decrease the gap of the communication. LSTM, bidirectional RNN, GRU, and CNN are mostly used models for recognition, sign language, and conversion. [12]. This paper introduced a deep-learning CNN model to recognize static signs. A totalof 35000 images are collected from different persons for different 100 static signs [13]. This research paper shows a deep learning algorithm, 5-layer CNN model proposed. The dataset of images of ISL was collected in simple and complex backgrounds [14]. Feature extraction is a preprocessing step that involves morphological filters, segmentation, contour generation, approximation, and polygonal. Testing and training are performed with different CNNs [15]. In this research work of Sruthi, proposed a method of deep learning that recognized Indian sign language (ISL) alphabets. It uses a CNN architecture to recognize ISL alphabets founded on the binary silhouette of the signer's hand region [16].

Table 1 describes the summary of sign language recognition with different parameters like Sign Language, Data collection device, Data acquisition, Algorithm used, and recognition rate of different research papers. From the above data, no papers provide the recognition of Gujarati Sign Language one paper did research in Hindi Sign Language as it provides 74% accuracy only. Most papers represent the English alphabet recognition in American Sign Language or Indian Sign Language. One of the research papers [28] provides a 100% recognition rate for Indian Sign Language alphabets. In Sign Language recognition uses a deep learning approach like LSTM, CNN, RNN, and SVM. The authors used web cameras to capture images as dataset collection, and some of the authors also used a leap motion sensor to capture signs.

3. METHODOLOGY

A. System Overview

The proposed structure working is depicted in Figure 1. The process of the proposed system is divided into four stages. These are i) data acquisition, ii) image pre-processing, iii) feature extraction, and iv) sign recognition. The whole proposed system is explained in the following steps:

Step 1: Take images of different people and different angles. Around 20 images of a particular sign were collected.

Step 2: The data augmentation technique is applied to increase the sample size because a small dataset size is not enough to train the system.

Step 3: Data augmentations: various functions are used to convert 1 image to 50 images to train.

Step 4: Deep learning algorithm Long-Short term A network is applied to train the model.

Step 5: It converts each image into a numpy array, which understands machines.

Step 6: Training the model with numpy array files, which are plucking out images.

Step 7: Train the model with 80% of the data, and the other 20% of the datais used for testing.

Step 8: To check the model, Input a sign in a camera and check its output.

Step 9: If the desired accuracy is not achieved, then train the model again.

Step 10: If desired accuracy and recognition of a correct alphabet are achieved, then check for other Gujarati alphabets also.

The system recognizes the signs of the Gujarati alphabet in real-time using a camera.

B. Data acquisition

The Gujarati Sign Language alphabet dataset was prepared with a collection of images. The Gujarati Sign Language dataset is not available digitally. As there is a need to take help from deaf-mute school educators and students to collect the sign or images. The dataset images were collected from distinct people. The dataset includes diverse RGB images taken from diverse positions and angles, featuring varying backgrounds of light and dark combinations. The dataset size was also not adequate for the model. The augmentation technique is applied to increase the datasetsize. This sample model is tested for 6 alphabets of Gujarati Sign Language like “ક ખગ ઘ ચ છ”. There are some example images of the “ખ’ alphabet in Figure 2 from the original dataset.

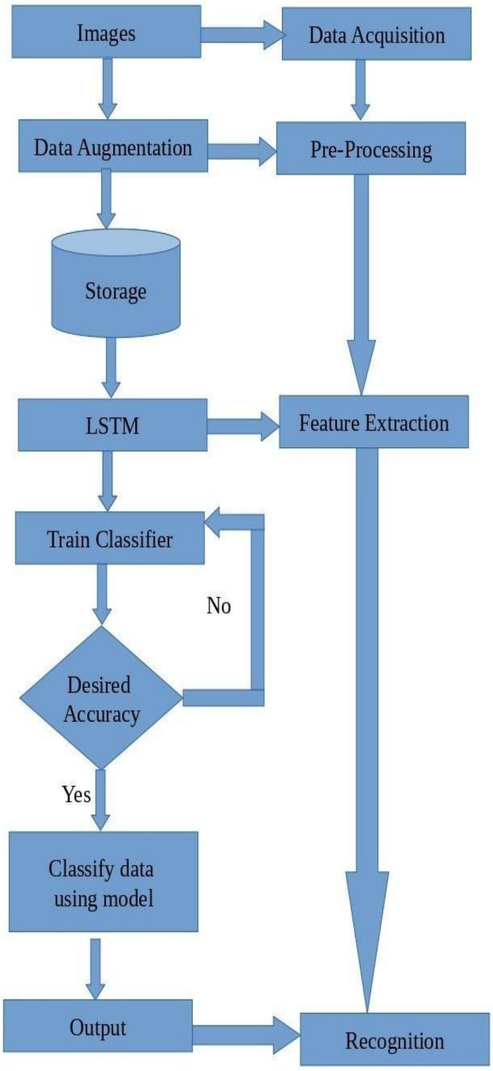

Figure 1. Flow Chart of Proposed System

C. Image Pre-processing

Data augmentation is a technique that is used to increase the size of your dataset using existing data to create a modified copy of the dataset. The original Dataset has 20 images per sign. Data augmentation was applied on input images to provide training so that the model cannot overfit. We took 1 image and created 50 images through augmentation like ColorJitter contrast, Color Jitter(brightness), CLAHE, Channel Shuffle, Random Gamma, RGBShift, etc. methods. So, we created 50 images for each image per sign = 1000 images, and for 5 images generated 5000 images. It used 80% applicable to training and the rest for validation. Figure 3 shows the augmented data. First, it converts data to constant resolution, All augmented images were taken and converted to 500*700 pixels to standard format for all images. Hand landmark detection is a pivotal component of computer vision, widely applied in diverse fields such as gesture recognition, and sign language interpretation. In this study, we utilized the Media Pipe library, coupled with OpenCV for image processing, to implement an effective hand landmark detection system to collect data for recognition of sign language. The methodology involved initializing the Media

Pipe Hand module, capturing frames from an image, and processing these frames to detect hand landmarks. We used only the right hand for capturing all signs so we made the system in such a way that it only detects the right hand in the camera and collects hand key points. It Stores landmarks as a NumPy array, each hand consists of 23 landmarks. Each landmark provides x, y, and z coordinates for a total of 63. The x and y coordinates are normalized before being stored as 1D NumPy arrays of size 63.

Figure 2. Sample Original dataset of “ખ ” Alphabet

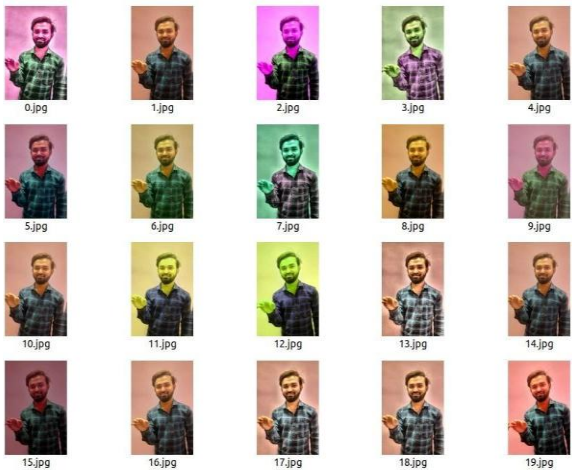

Figure 3. Data Augmentation dataset of “ખ ” Alphabet

D. Feature Extraction

In this study, authors used an LSTM model with 3 layers and then processed through a linear layer to train the model on 1D NumPy arrays. We chose LSTM over CNN since our model does not analyze images directly, but rather the extracted holistic key points. The output dense layer consists of 5 units with Soft-Max activation. The model is trained for a thousand epochs with a batch size of 8 with Adam optimization. We trained with 4000 images and tested with 1000 images. We achieved over 90% accuracy on our test set with this small sample of five signs. To train the model, the Nvidia RTX 3070 8 GB GPU was used.

E. Image recognition

Image recognition is a final step used by authors to recognize Gujarati sign language. The authors used Pytorch and OpenCV python libraries for data loading and reading images from the webcam. OpenCV shot real-time hand-shaped videoframes from the user. The system successfully detects and predicts sign alphabets. The proposed structure successfully predicts the Gujarati Sign Language alphabet shown in Figures 4.1, 4.2, 4.3, and 4.4. Alphabets of Gujarati Sign Language Recognition. Figure 4.1 shows the “ક” alphabet recognized by a person. A person needs to sign in a web camera through hands. According to the model, it recognizes the sign with its accuracy if it shows correct hand gestures. In the model trained by the right hand, the Person needs to sign with the right hand only then does it show the alphabet recognition output. If a person does a sign by left hand it will not recognize the alphabet. This whole process is the same for other alphabets. Let's summarize the process of the whole model. First, we need to collect images with different people, different locations, backgrounds, angles, and distances with particular alphabet signs. For eg. If we collect 20 images per sign then we augment the images to train the model. 20 images are not enough to train the model, so we converted images with augmentation techniques with blur, contrast, resize, angle, and so on to increase the size of the sample data. 80% data is used to prepare the model and 20% data is used to check the model. It gives 98% accuracy score to identify the alphabet. The above images are examples of the "ક", "ખ", "ગ", and "ઘ".

Figure 4.1. Sign Recognition of “ક ” Alphabet

Figure 4.2. Sign Recognition of “ખ ” Alphabet

Figure 4.3. Sign Recognition of “ગ ” Alphabet

Figure 4.4. Sign Recognition of “ઘ ” Alphabet

4. Result Analysis

The above model provides 98% accuracy as it recognizes the Gujarati alphabet correctly. This model uses an LSTM algorithm to train the model. Other authors also use different deep learning algorithms such as CNN, SVM, and LSTM to recognize the different sign languages. One of the authors recognized the Hindi alphabets of Indian Sign Language as they are the same as Gujarati Sign Language alphabets. It provides 74% accuracy only using the CNN algorithm. To find out our model accuracy, we calculate precision, recall, and f1 score. Other authors also recognized American sign language with different deep-learning algorithms. American sign language with the CNN algorithm provides 98.98% accuracy and with the SVM algorithm, it provides 99% accuracy.

Precision: Precision is the accuracy of the positive prediction from the total positive prediction. A high precision means it predicts a certain alphabet almost correctly.

Precision = True positives/ (True positives + False positives)

Recall: Recall is also known as true positive rate, It is a ratio of correctly positive prediction observations to all actual positives. It shows well the model can find all the relevant instances within a dataset. A high recall suggests that the model is good at identifying all instances of a particular alphabet.

Recall = True Positive (TP) / True Positive (TP) + False Negative (FN)

F1-score:The F1-score is the harmonic mean of precision and recall-;. It's a single metric that combines both precision and recall into one value, providing a balance between the two. It calculated from the formula

F1score = 2x [(PrecisionxRecall) / (Precision+Recall)]

Table 1 : Summary of Accuracy Result

Sign	precision	recall	f1_score
ચ	1.0	1.0	1.0
છ	1.0	0.99	0.99
ગ	0.98	0.93	0.95
ઘ	0.96	0.98	0.97
ક	0.97	1.0	0.98
ખ	0.99	0.98	0.99

Alphabet: This column lists different alphabets(e.g., ક ખગ ઘ ચ છ)that the model is predicting.
Precision: This column shows the precision score for each alphabet. Figure 5 shows the Accuracy Test. For instance: The class 'ચ' has a precision of 1.0, indicating that all instances predicted as 'ચ' by the model are true 'cha'. The class 'ગ' has a precision of 0.98, meaning that when the model predicts 'ગ', it's correct 98% of the time.
Recall: This column shows the recall score for each alphabet. For example, The class 'ચ' has a recall of 1.0, meaning the model correctly identifies all actual instances of 'ચ'. The class 'ગ' has a recall of 0.93, suggesting that the model identifies only 93% of actual instances of 'ગ'.
F1-score: This column presents the F1 score for each alphabet, calculated using the precision and recall values. For instance: The class 'cha' has an F1-score of 1.0, which is the harmonic mean of precision (1.0) and recall (1.0). The class 'ga' has an F1-score of 0.95, which is a balanced measure of precision (0.98) and recall (0.93).

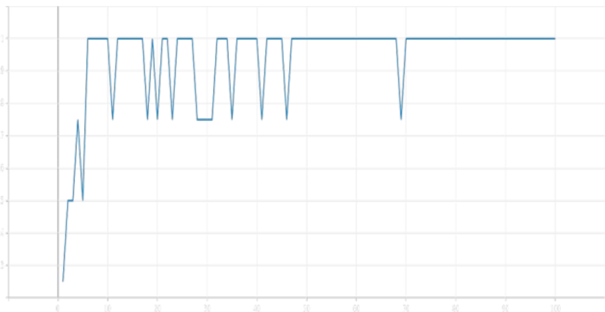

Figure. 5. Accuracy Test

5. CONCLUSION

The main goal of this research idea was to identify the Sign Language alphabets of Gujarati using a webcam. There were many developments done in various sign language recognition. This is the development for Gujarati Sign Language recognition using a camera that can be easily accessible for everyone. This developed system can also used in deaf dumb schools of Gujarat to teach and test alphabets to students. For this study authors first created the Gujarati Sign Language Alphabet dataset and augmented it to train the model. The authors did data augmentation to convert 1 alphabet into 20 images of signs of different persons and augmented 1 image = 50 images, so per alphabet sign has 1000 images to train the model, 80% used to train the model and 20% used for validation. A deep learning-based LSTM model is used to train the model. Result Analysis shows that each alphabet precision, recall, and f1 score value. Almost it gives 98% accuracy of each alphabet. This study will help to decrease the communication gap between deaf-mute and normal people.

6. FUTURE ENHANCEMENT

Sign Language is widely used in the deaf community. The goal of the research is to provide a system with all Gujarati consonants, vowels, and mantras to finger-spell all words. The given experiment was used to detect and recognize individual and isolated signs of consonants. This can be further modified to recognize continuous sign language for more practical use. In a future enhancement develop vowels and matras. Any deaf person can perform signs into a system that gives an output as a text or speech. So, any other person can easily understand the conversation of a deaf person.

REFERENCES

[1] Ibrahim, Nada B., Mazen M. Selim, and Hala H. Zayed. "An automatic Arabic sign language recognition system (ArSLRS)." Journal of King Saud University-Computer and Information Sciences 30.4 (2018): 470-477.

[2] Wangchuk, Karma, Panomkhawn Riyamongkol, and Rattapoom Waranusast. "Real-time Bhutanese sign language digits recognition system using convolutional neural network." Ict Express 7.2 (2021): 215-220.

[3] Adeyanju, I. A., O. O. Bello, and M. A. Adegboye. "Machine learning methods for sign language recognition: A critical review and analysis." Intelligent Systems with Applications 12 (2021): 200056.

[4] Lee, Carman KM, et al. "American sign language recognition and training method with recurrent neural network." Expert Systems with Applications 167 (2021): 114403.

[5] Gupta, Pratibha, et al. "SIGN LANGUAGE RECOGNITION FOR HINDI VARNAMALA USING CNN.".

[6] Amrutha, K., and P. Prabu. "ML based sign language recognition system." 2021 International Conference on Innovative Trends in Information Technology (ICITIIT). IEEE, 2021.

[7] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage: www. ijrpr. com ISSN 2582 (2021): 7421.

[8] Pathan, Refat Khan, et al. "Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network." Scientific Reports 13.1 (2023): 16975.

[9] Rao, G. Mallikarjuna, et al. "Sign Language Recognition using LSTM and Media Pipe." 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2023.

[10] Grover, Chhaya, et al. "Sign Language Recognition in the Hindi Language Based on Computer Vision." Authorea Preprints (2023).

[11] Wadhawan, Ankita, and Parteek Kumar. "Sign language recognition systems: A decade systematic literature review." Archives of Computational Methods in Engineering 28 (2021): 785-813.

[12] González-Rodríguez, Jaime-Rodrigo, et al. "Towards a Bidirectional Mexican Sign Language–Spanish Translation System: A Deep Learning Approach." Technologies 12.1 (2024): 7.

[13] Wadhawan, Ankita, and Parteek Kumar. "Deep learning-based sign language recognition system for static signs." Neural computing and applications 32 (2020): 7957-7968.

[14] Dhiman, R., G. Joshi, and C. Rama Krishna. "A deep learning approach for Indian sign language gestures classification with different backgrounds." Journal of Physics: Conference Series. Vol. 1950. No. 1. IOP Publishing, 2021.

[15] Pinto, Raimundo F., et al. "Static hand gesture recognition based on convolutional neural networks." Journal of Electrical and Computer Engineering 2019 (2019): 1-12.

[16] Sruthi, C. J., and A. Lijiya. "Signet: A deep learning based indian sign language recognition system." 2019 International conference on communication and signal processing (ICCSP). IEEE, 2019.

[17] Aasofwala, Nasrin, Shanti Verma, and Kalyani Patel. "A Novel Speech to Sign Communication Model for Gujarati Language." 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2021.

[18] Aasofwala, Nasrin, Shanti Verma, and Kalyani Patel. "NLP based model to convert English speech to Gujarati text for deaf & dumb people." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[19] Aasofwala, Nasrin, Shanti Verma, and Kalyani Patel. "Conversion of Gujarati Alphabet to Gujarati Sign Language Using Synthetic Animation." International Conference on ICT for Sustainable Development. Singapore: Springer Nature Singapore, 2023.

[20] Aasofwala, Nasrin, and Shanti Verma. "SURVEY ON GUJARAT'S DEAF AND MUTE SCHOOL'S SPECIAL EDUCATORS TO UNDERSTAND THE EDUCATION OF DEAF COMMUNITY IN THIS PANDEMIC." Towards Excellence 13.4 (2021).

[21] Patel, Dhaval U., and Jay M. Joshi. "Deep Leaning Based Static Indian-Gujarati Sign Language Gesture Recognition." SN Computer Science 3.5 (2022): 380.

[22] Tolentino, Lean Karlo S., et al. "Static sign language recognition using deep learning." International Journal of Machine Learning and Computing 9.6 (2019): 821-827.

[23] Bantupalli, Kshitij, and Ying Xie. "American sign language recognition using machine learning and computer vision." (2019).

[24] Kothadiya, Deep, et al. "Deepsign: Sign language detection and recognition using deep learning." Electronics 11.11 (2022): 1780.

[25] Shin, Jungpil, et al. "American sign language alphabet recognition by extracting feature from hand pose estimation." Sensors 21.17 (2021): 5856.

[26] Raheja, J. L., Anand Mishra, and Ankit Chaudhary. "Indian sign language recognition using SVM." Pattern Recognition and Image Analysis 26 (2016): 434-441.
[27] Chuan, Ching-Hua, Eric Regina, and Caroline Guardino. "American sign language recognition using leap motion sensor." 2014 13th International Conference on Machine Learning and Applications. IEEE, 2014.
[28] Shirbhate, Radha S., et al. "Sign language recognition using machine learning algorithm." International Research Journal of Engineering and Technology (IRJET) 7.03 (2020): 2122-2125.