Journal of Autonomous Intelligence (2023) Volume 6 Issue 3
doi: 10.32629/jai.v6i3.1000

ORIGINAL RESEARCH ARTICLE

Indian sign language recognition and search results

Sandeep Musale*, Kalyani Gargate, Vaishnavi Gulavani, Samruddhi Kadam, Shweta Kothawade

Department of Electronics and Telecommunication, Cummins College of Engineering for Women, Pune 411052, India

*** Corresponding author:** Sandeep Musale, sandeep.musale@cumminscollege.in

ABSTRACT

Sign language is a medium of communication for people with hearing and speaking impairment. It uses gestures to convey messages. The proposed system focuses on using sign language in search engines and helping specially-abled people get the information they are looking for. Here, we are using Marathi sign language. Translation systems for Indian sign languages are not much simple and popular as American sign language. Marathi language consists of words with individual letters formed of two letter = Swara + Vyanjan (Mulakshar). Every Vyanjan or Swara individually has a unique sign which can be represented as image or video with still frames. Any letter formed of both Swara and Vyanjan is represented with hand gesture signing the Vyanjan as above and with movement of signed gesture in shape of Swara in Devnagari script. Such letters are represented with videos containing motion and frames in particular sequence. Further the predicted term can be searched on google using the sign search. The proposed system includes three important steps: 1) hand detection; 2) sign recognition using neural networks; 3) fetching search results. Overall, the system has great potential to help individuals with hearing and speaking impairment to access information on the internet through the use of sign language. It is a promising application of machine learning and deep learning techniques.

Keywords: Swar; Vyanjan; Mulakshar; Devnagari; neural networks; Indian sign language

ARTICLE INFO

Received: 18 July 2023
Accepted: 31 July 2023
Available online: 22 August 2023

1. Introduction

Sign language is an essential mode of communication for people who have speaking or hearing disabilities, but it is not widely used by others. Sign language recognition can help bridge the communication gap between the disabled and other people. Additionally, a tool that searches using sign language has been included for those who struggle with speech clarity and elderly persons who tend to type slowly. The proposed system aims to capture images of hand gestures signing Marathi letters and convert them to equivalent text and feed it to a search engine to retrieve related search results. The Marathi language consists of words with individual letters formed of two parts letter = Swara + Vyanjan (Mulakshar). We are using video preprocessing[1,2] for Swara determination and image classification[3,4] for Vyanjan determination. Then we are concatenating the Vyanjan with Swara to determine the letters. Further, the predicted term can be searched on Google using the sign search.

The proposed system emphasizes studying various methods and technologies for sign language recognition and implementing one to model a basic level of sign language support on search engines.

Objective

The proposed system aims to take input as hand gesture images and videos of signing Marathi letters and convert it to text, feeding the text to a search engine to get search results. Various Marathi sign language mulakshar is shown in **Figure 1**.

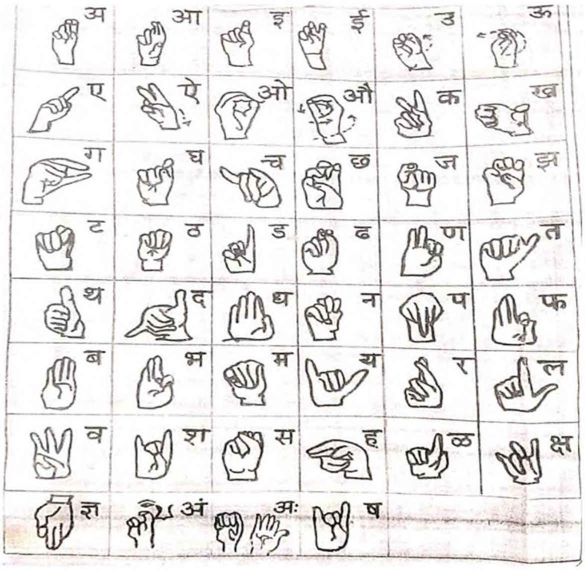


Figure 1. Marathi sign language Mulakshar.

2. Literature survey

The analysis of literature encompasses various research papers on sign language recognition, in which various machine learning algorithms are used to detect sign language from images and videos.

Patil et al.[5] details about a series of processing steps which include various computer vision techniques such as the conversion to gray-scale, dilation and mask operation are covered. CNN is used to train the model and identify the pictures. Rani et al.[6] includes gesture classification which contains preprocessing steps and CNN is used as classifier with 2 layers on American sign language (ASL). The challenge encountered was the requirement of square images for CNN implementation in Keras. Sood[7] gives us insight about the technique called transfer learning in combination with data augmentation. Google's inception v3 model (image recognition model) is implemented in this paper. In the study of Haldera and Tayadeb[3], pre-processing of images using mediapipe is done to get multi-hand landmarks. Machine learning algorithms and deep learning algorithms like SVM, ANN, KNN, etc., are implemented and compared. Darekar et al.[8] covers Marathi sign language recognition with random forests algorithm, and classifier methods. Overview of entire methodology and image pre-processing is explained. Katoch et al.[1] covers SVM and CNN technique used in combination for sign language recognition and compares their performance metrics. Shinde and Kagalkar[9] covers hidden markov model (HMM), euclidean distance, sensor based, proposed system. In the study of Mali et al.[4], machine learning model is created using SVM classifier. In this paper accuracy ranges 90%–96.87% when proposed by different people. Subramanian et al.[10] covers methodology using mediapipe for feature extraction and GRU model is used for RNN hand gesture recognition.

3. Proposed methodology

3.1. Dataset creation

Indian (Marathi) sign language consists of signs with hand gestures both stationary as well as in motion. Thus, it is required to create a dataset of images as well as videos for effective recognition. The schematic of the proposed methodology is shown in **Figure 2**.

1) Image dataset creation

The image dataset includes images of a total 7 Vyanjans in Marathi with around 120 images of each Vyanjan. As the number of categories are more, the dataset needs to be expanded as required. The images of hand gestures are captured through a webcam of resolution 360p and added in respective Vyanjan folders named according to the English phonetics (used in python Indic transliteration library) of the Marathi Vyanjan. Data augmentation[11] is performed to expand the dataset that includes rotation, flipping and adding noise to captured images.

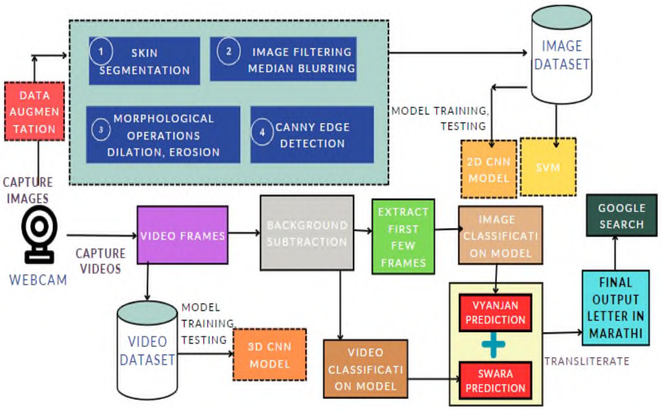


Figure 2. Block schematic diagram.

2) Video dataset creation

The video dataset includes videos of motion of hands signing 12 different Swaras in Marathi barakhadi. For each of the 6 Swaras, motions of hand poses signing compound letters of form Vyanjan + Swara, are captured by changing Vyanjan but having same Swara. This is done for all Swaras. The videos are captured through webcam of minimum 30 frames per second rate and every 2nd frame is stored in the dataset that contains maximum 90 frames per video.

3.2. Image preprocessing and classification for Vyanjan recognition

1) Image preprocessing: The images are preprocessed[5]. Before being used for training for efficient feature extraction to detect the shape of the hand gesture signing a particular Vyanjan[1,12].

a) Median blurring: Median blurring is an image filtering technique performed for reducing noise. The median of 3 × 3 kernel size of image replaces the center pixel value of the image portion covered by the kernel.

b) Morphological operations: These techniques are effectively used for sign language recognition. These operations help to enhance shape of the hand pose in the image. Dilation of image is done as mentioned in **Figure 3** followed by erosion as mentioned in **Figure 4** for image enhancement. The structuring element S is a kernel of white pixels of dimensions 5 × 5.

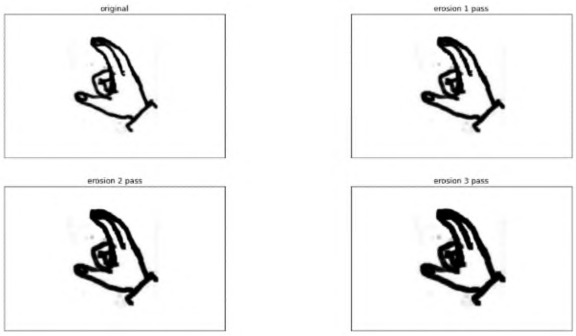


Figure 3. Dilation: I ⊕ S.

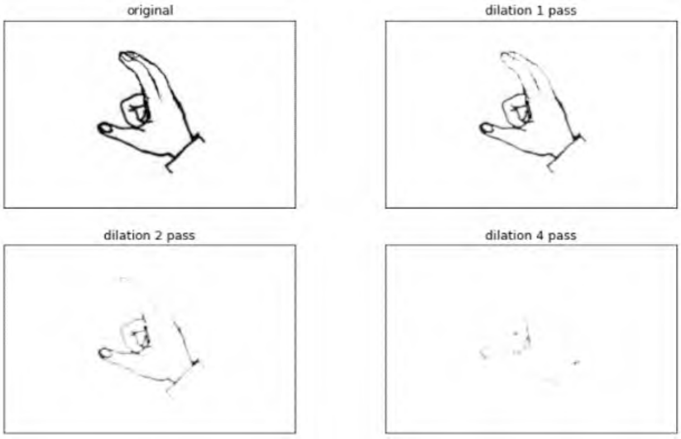


Figure 4. Erosion: $I \ominus S$.

c) Skin segmentation: skin segmentation algorithm[12] is implemented to segment hand gesture and eliminate background from input image and convert it to a binary image using special relationship amongst RGB values of skin colour. The pixel colour can be identified as skin colour if the RGB values satisfy the relationship r > 90 and g > 40 and b > 20 and (max (RGB)-min (RGB) > 15) and r > g and r > b and abs (r-g) > 15. The segmented images are binary images of size 100 × 100.

d) Canny edge detection: canny edge detection is used for detecting edges in the image and the edge detected image is given as input to deep learning model for training. The kernels are

x-direction kernel

−1	0	1
−2	0	2
−1	0	1

y-direction kernel

−1	−2	−1
0	0	0
1	2	1

2) Model building and training: the above preprocessed and edge detected images are stored as intermediate data and used for training the deep learning recognition models for Vyanjan recognition. Sign language recognition using images is performed with SVM[1] and CNN[1,2,13] the two deep learning techniques under consideration are:

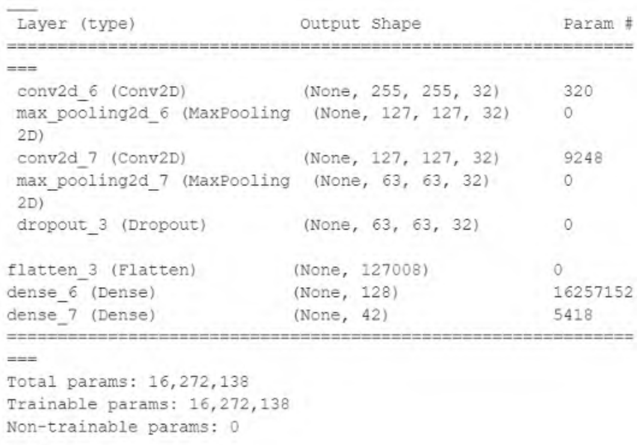
a) Support vector machine:

Support vector machines (SVM) is used for image classification for sign language recognition[3,4]. Herein, it is used for multiclass classification of image dataset for Vyanjan recognition. Support vector classification SVC uses one-vs-one approach for classification implementing n × (n − 1)/2 classifiers each classifying data of two training classes. The binary SVC tries to solve the problem

$$min_{\omega,b}\frac{1}{2}\omega^{T}\omega + C\sum_{i=1}^{n} max(0, 1 - y_i(\omega^{T}\emptyset(x_i) + b))$$

SVC herein uses linear kernel function for classification.

b) 2D convolutional neural networks:



```
___
 Layer (type)                Output Shape              Param #
=================================================================
===
 conv2d_6 (Conv2D)           (None, 255, 255, 32)      320
 max_pooling2d_6 (MaxPooling  (None, 127, 127, 32)     0
 2D)
 conv2d_7 (Conv2D)           (None, 127, 127, 32)      9248
 max_pooling2d_7 (MaxPooling  (None, 63, 63, 32)       0
 2D)
 dropout_3 (Dropout)         (None, 63, 63, 32)        0

flatten_3 (Flatten)         (None, 127008)            0
dense_6 (Dense)             (None, 128)               16257152
dense_7 (Dense)             (None, 42)                5418
=================================================================
===
Total params: 16,272,138
Trainable params: 16,272,138
Non-trainable params: 0
```

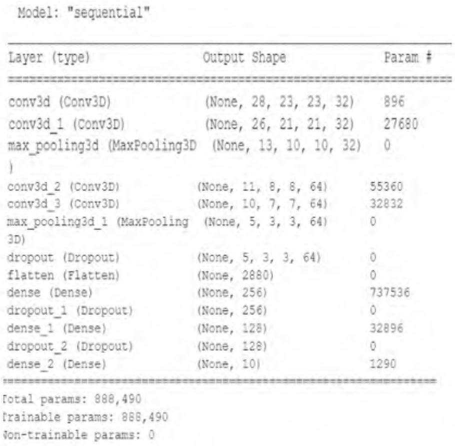
Figure 5. 2D CNN training of the image dataset.

Convolutional neural networks are used for image classification. The input is applied to convolutional layer of 32 filters and 3 × 3 kernel size and uses padding along with ReLu activation function. This layer is followed by pooling layer of with kernel size 2 × 2 and strides =2. Similar set of layers is repeated again followed by dropout layer 0.5 as shown in **Figure 5**. In the next set the filters are reduced to 16. The layers are flattened and a dense layer of 128 output classes is implemented with ReLu activation function followed by the final output layer of 33 classes with softmax activation function.

3.3. Video classification for Swara recognition

We have used 25 videos per class for video classification. There are total 6 no. of classes. The videos from captured dataset are pre-processed to get uniform dimensions (90, 25, 25, 1) where 90 is maximum frames per video and 25 × 25 is the size of the grayscale frame (number of channels =1). Classification is done through 3D convolutional neural networks[1,2].

The 3D CNN architecture has input convolutional layer of 32 filters, kernel size 3 × 3 × 3, ReLu activation function and bias initializer as 0.01. The same layer is repeated followed by MaxPooling layer of size 2 × 2 × 2. Next the number of filters in convolutional layers are reduced to 64 and size of the kernel is 3 × 3 × 3 followed by next layer with kernel size 2 × 2 × 2. One more MaxPooling layer is added of size 2 × 2 × 2 and dropout of 0.6 followed by final 3D convolutional layer with 128 filters and kernel size 3 × 3 × 3. All convolutional layers above use ReLu activation. The layers are finally flattened and then dense layers with ReLu activation having output classes 256, dropout 0.7 and another one with output classes 128, dropout 0.5 are added. The final dense output layer has 12 output classes representing each Swara and uses softmax activation.



```
 Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv3d (Conv3D)             (None, 28, 23, 23, 32)    896
 conv3d_1 (Conv3D)           (None, 26, 21, 21, 32)    27680
 max_pooling3d (MaxPooling3D  (None, 13, 10, 10, 32)   0
 )
 conv3d_2 (Conv3D)           (None, 11, 8, 8, 64)      55360
 conv3d_3 (Conv3D)           (None, 10, 7, 7, 64)      32832
 max_pooling3d_1 (MaxPooling  (None, 5, 3, 3, 64)      0
 3D)
 dropout (Dropout)           (None, 5, 3, 3, 64)       0
 flatten (Flatten)           (None, 2880)              0
 dense (Dense)               (None, 256)               737536
 dropout_1 (Dropout)         (None, 256)               0
 dense_1 (Dense)             (None, 128)               32896
 dropout_2 (Dropout)         (None, 128)               0
 dense_2 (Dense)             (None, 10)                1290
=================================================================
Total params: 888,490
Trainable params: 888,490
Non-trainable params: 0
```

Figure 6. 3D CNN training of the video dataset.

3.4. Marathi letter prediction on video input data

The recognition is to be performed on video input data signing a compound Marathi letter of form Vyanjan + Swara. Prediction of Vyanjan and Swara is done separately and then both are concatenated to predict the final compound letter.

1) Vyanjan prediction: for predicting Vyanjan, every 2nd frame of input video is preprocessed and given as input to either SVM or CNN image classification model. Predictions for 10 frames are performed and mode of the predictions is considered as the final Vyanjan prediction for the input video.

2) Swara prediction: for Swara prediction, background subtraction algorithm[13] is applied on input video to track changes in consecutive video frames that check if the hand gesture in video is stationary or in motion. MoG background subtraction method is chosen wherein pixel locations of background are represented as Gaussian probability distributions as

$$\mathrm{k}$$

$$\mathrm{F}\ (\mathrm{i}_{\mathrm{t}} = \mu)\ \sum_{i=-1}^{k} \omega_{i,t} \cdot \eta\ (\mu, \sigma)$$

η = i-th gaussian component,
$\omega_{i,t}$ = portion of data accounted by i-th componenet,
μ = intensity mean,
σ = satndard deviation.

The presence of a foreground image indicates a frame in motion and absence of a foreground object indicates a stationary frame. Every alternate non-stationary frame is stored and the Swara prediction is done based on the presence of stationary or non-stationary frames as follows:

a) If the number of non-stationary frames in the input video is less than 10, it implies that the hand gesture in the input video is stationary and the signed Marathi letter is a pure Vyanjan with Swara “a”.

b) If the input video contains non-stationary frames, every 2nd frame non-stationary frame is stored and maximum 90 frames per video with dimensions 25 × 25 are given to 3D CNN video classification model for Swara prediction.

3) Output Marathi letter prediction: the Vyanjan and Swara predicted are concatenated together to form the compound letter predicted of form Vyanjan + Swara which is then translated to Marathi output using the Indic transliteration library in python. The output compound letter is displayed in Marathi. Streams of such letters can be predicted to recognize a signed word or sentence.

3.5. Output

Concatenating Vyanjan with Swara to determine letters: once the Vyanjan and Swara are identified, they are concatenated to form letters. In Marathi, a letter is made up of a Vyanjan and a Swara. By concatenating the Vyanjan and Swara, the letters in the video can be determined.

Using transliteration in python to display output in Marathi: to display the output in Marathi, transliteration can be used. Transliteration involves converting the text from one script to another. In this case, the output text in English can be transliterated into the Marathi script using python libraries such as Indic transliteration. The output can then be displayed in the Marathi script for easy readability.

3.6. Implementing sign search

The predicted Marathi output is fed to Google search engine through API key to retrieve search results of the signed term in Marathi.

3.7. Performance evaluation

Performance parameters of the neural network is evaluated as follows:

Confusion matrix is plotted for the model with 33 output classes that gives a clear picture of all TP, FP, TN and FN of all the classes. These values help to determine the following performance evaluation parameters:

1) Precision: TP/(TP + FP)

It is the ratio of true positives to all the predicted positive values.

2) Accuracy: (TP + TN)/No. of test samples

It is the ratio of correctly predicted values to the total number of samples.

3) Recall: TP/(TP + FN)

It is the ratio of truly predicted positives to the actual number of positive samples.

4) F1-score: 2 × (Precision × Recall)/(Precision + Recall)

It is the harmonic mean of precision and recall.

4. Result

Table 1. Classification report of SVM.

	Precision	**Recall**	**F1-score**	**Support**
0	0.82	0.75	0.78	24
1	0.88	0.92	0.90	24
2	0.75	0.88	0.81	24
3	1.00	0.71	0.83	24
4	1.00	1.00	1.00	24
5	0.96	0.92	0.94	24
6	0.66	0.79	0.72	24
-	-	-	-	-
Accuracy	-	-	0.85	168
Macro average	0.87	0.85	0.85	168
Weighted average	0.87	0.85	0.85	168

Table 2. Classification report of CNN.

	Precision	**Recall**	**F1-score**	**Support**
0	0.91	0.84	0.87	25
1	1.00	0.96	0.98	27
2	1.00	0.95	0.98	21
3	0.83	0.95	0.89	21
4	1.00	1.00	1.00	22
5	1.00	1.00	1.00	24
6	0.90	0.93	0.91	28
-	-	-	-	-
Accuracy	-	-	0.95	168
Macro average	0.95	0.95	0.95	168
Weighted average	0.95	0.95	0.95	168

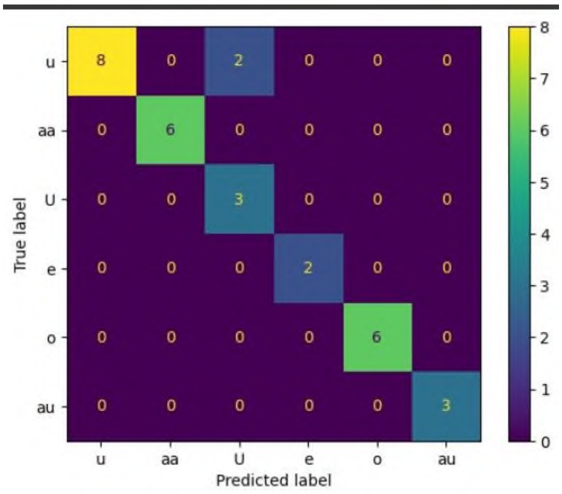
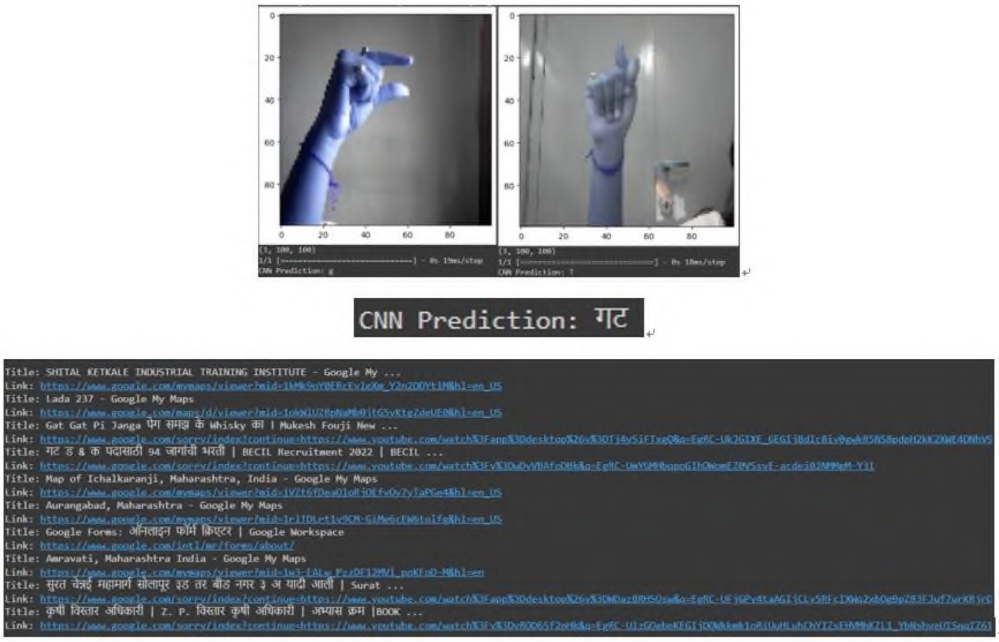


Figure 7. 3D CNN Confusion matrix.



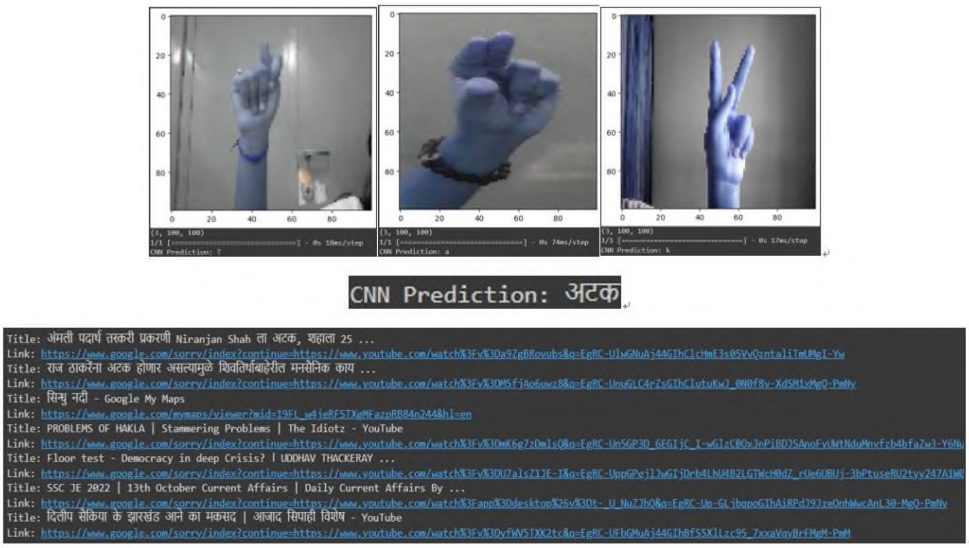
```
(3, 100, 100)
1/1 [==============================] - 0s 19ms/step
CNN Prediction: g
```

```
(3, 100, 100)
1/1 [==============================] - 0s 18ms/step
CNN Prediction: T
```

```
CNN Prediction: गट
```

```
Title: SHITAL KETKALE INDUSTRIAL TRAINING INSTITUTE - Google My ...
Link: https://www.google.com/mymaps/viewer?mid=1kMkSoY0ERcEvlzXm_Y2n2ODYt1M&hl=en_US
Title: Lada 237 - Google My Maps
Link: https://www.google.com/maps/d/viewer?mid=1okWlUZRpNaMb0jtG5yKtg7deUE0&hl=en_US
Title: Gat Gat Pi Janga पेग समझ के Whisky का | Mukesh Fouji New ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fapp%3Ddesktop%26v%3DTj4ySiFTxgQ&q=EgRC-UkJGIXE_GEGIj8dIc8iy0gwk8SNS8pdpH2kK2XWE4DNhVS
Title: गट ड & क पदासाठी 94 जागांची भरती | BECIL Recruitment 2022 | BECIL ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DwDyVBAfoD0k&q=EgRC-UwYGMHbqpoGIhOWomEZ0VSsvE-acdei02NMMgM-Y1I
Title: Map of Ichalkaranji, Maharashtra, India - Google My Maps
Link: https://www.google.com/mymaps/viewer?mid=1VZt6fOea0loRjOEfvOy7yTaPGe4&hl=en_US
Title: Aurangabad, Maharashtra - Google My Maps
Link: https://www.google.com/mymaps/viewer?mid=1rlIDLrt1y9CM-GiMe6cEW6tolfe&hl=en_US
Title: Google Forms: ऑनलाइन फॉर्म क्रिएटर | Google Workspace
Link: https://www.google.com/intl/mr/forms/about/
Title: Amravati, Maharashtra India - Google My Maps
Link: https://www.google.com/mymaps/viewer?mid=1w3-EALw_PzzOF12MVl_poKFoD-M&hl=en
Title: सुरत चेन्नई महामार्ग सोलापूर इड तर बीड नगर ३ अ यादी आली | Surat ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fapp%3Ddesktop%26v%3DWDazBRHS0sw&q=EgRC-UEjGPy4taAGIjCLy5RFcIXWq2xb0g0pZ83FJuf7urK8jrC
Title: कृषी विस्तार अधिकारी | Z. P. विस्तार कृषी अधिकारी | अभ्यास क्रम |BOOK ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DvR0D65f2pH8&q=EgRC-UlzG0ebeKEGIjDXWkkmk1oRiUuHLuhChYIZsEHMMhKZL1_YbNshveUTSgqIZ6I
```

Figure 8. Search results when the predicted word "Gat".



```
(3, 100, 100)
1/1 [==============================] - 0s 18ms/step
CNN Prediction: T
```

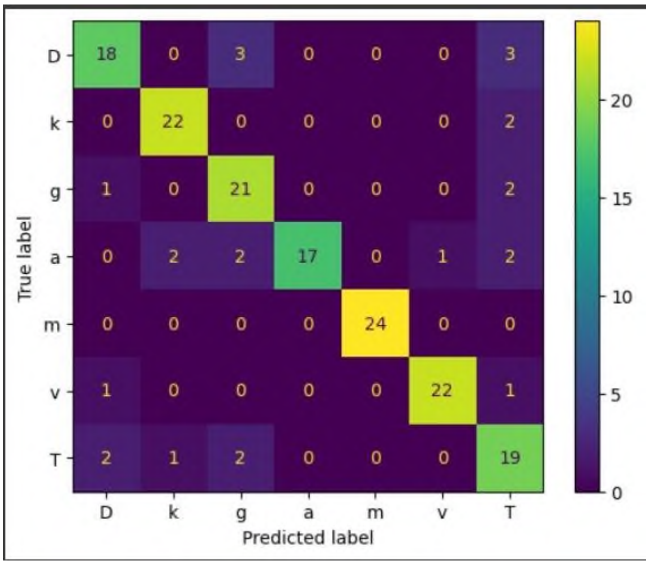
```
(3, 100, 100)
1/1 [==============================] - 0s 74ms/step
CNN Prediction: a
```

```
(3, 100, 100)
1/1 [==============================] - 0s 17ms/step
CNN Prediction: k
```

```
CNN Prediction: अटक
```

```
Title: अंमली पदार्थ तस्करी प्रकरणी Niranjan Shah ला अटक, शहाला 25 ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3Da97gBRovubs&q=EgRC-UlwGNuAj44GIhClcHmE3s05VvQzntaliTmUMgI-Yw
Title: राज ठाकरेंना अटक होणार असल्यामुळे शिवतिर्थाबाहेरील मनसैनिक काय ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DM5fjAo6uwz8&q=EgRC-UnuGLC4rZsGIhClutuKwJ_0N0f8y-XdSM1xMgQ-PmNy
Title: सिन्धु नदी - Google My Maps
Link: https://www.google.com/mymaps/viewer?mid=19FL_w4jeRFSTXgMEazpRB84n244&hl=en
Title: PROBLEMS OF HAKLA | Stammering Problems | The Idiotz - YouTube
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DmK6g7zDmlsQ&q=EgRC-UnSGP3D_6EGIjC_I-wGlzCBOxJnPiBDJSAnoFyUWtNduMnvfzb4bfaZw3-Y6Nu
Title: Floor test - Democracy in deep Crisis? | UDDHAV THACKERAY ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DU7alsZ1JE-I&q=EgRC-UppGPej1JwGIjDrb4LhU4B2LGTWcH0dZ_rUe6UBUj-3bPtuseRU2tyy247A1W0
Title: SSC JE 2022 | 13th October Current Affairs | Daily Current Affairs By ...
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fapp%3Ddesktop%26v%3Dt-_U_NuZJhQ&q=EgRC-Up-GLjbqpoGIhAiRPdJ9JzeOnhWwcAnL30-MgQ-PmNy
Title: दिलीप सैकिया के झारखंड आने का मकसद | आजाद सिपाही विशेष - YouTube
Link: https://www.google.com/sorry/index?continue=https://www.youtube.com/watch%3Fv%3DyfWV5TXK2tc&q=EgRC-UFbGMuAj44GIhBfSSXlLzc95_7xxaVqyBrFMgM-PmM
```

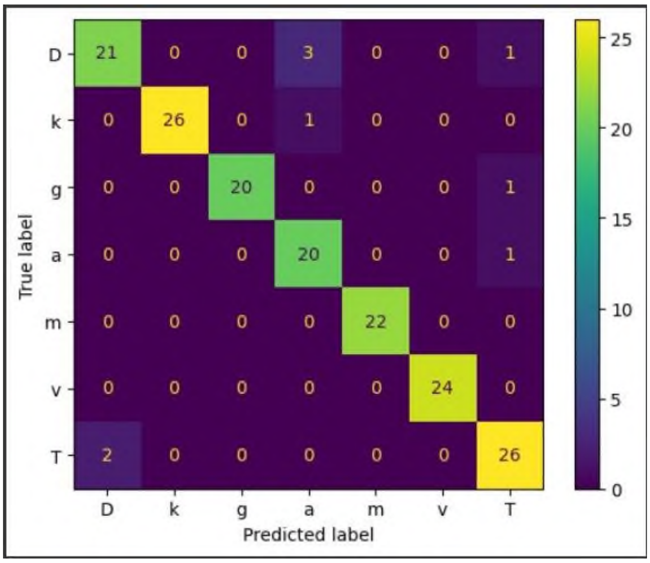
Figure 9. CNN prediction for "Mulakshar" for word "Atak" and search results when the predicted word "Atak".



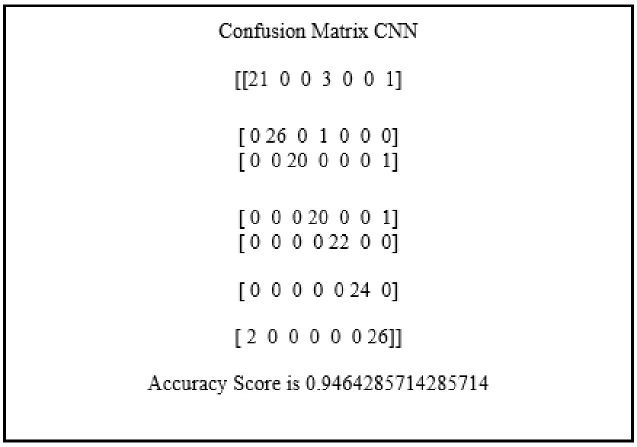
	D	k	g	a	m	v	T
D	18	0	3	0	0	0	3
k	0	22	0	0	0	0	2
g	1	0	21	0	0	0	2
a	0	2	2	17	0	1	2
m	0	0	0	0	24	0	0
v	1	0	0	0	0	22	1
T	2	1	2	0	0	0	19

```
Confusion Matrix for SVM:

[[18  0  3  0  0  0  3]

[ 0 22  0  0  0  0  2] 1  0 21  0  0  0  2]

[ 0  2  2 17  0  1  2]

[ 0  0  0  0 24  0  0]

[ 1  0  0  0  0 22  1]

[ 2  1  2  0  0  0 19]]

Accuracy Score is 0.8511904761904762
```

Figure 10. SVM confusion matrix.

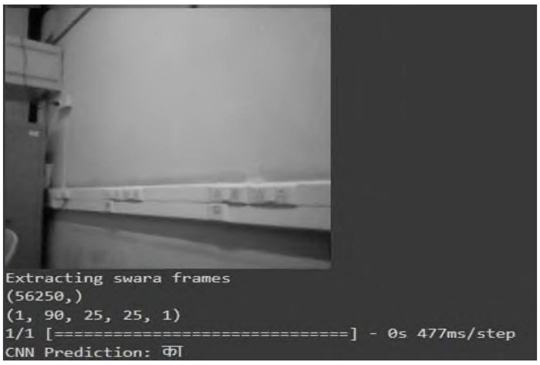


	D	k	g	a	m	v	T
D	21	0	0	3	0	0	1
k	0	26	0	1	0	0	0
g	0	0	20	0	0	0	1
a	0	0	0	20	0	0	1
m	0	0	0	0	22	0	0
v	0	0	0	0	0	24	0
T	2	0	0	0	0	0	26



```
Confusion Matrix CNN

[[21  0  0  3  0  0  1]

 [ 0 26  0  1  0  0  0]
 [ 0  0 20  0  0  0  1]

 [ 0  0  0 20  0  0  1]
 [ 0  0  0  0 22  0  0]

 [ 0  0  0  0  0 24  0]

 [ 2  0  0  0  0  0 26]]

Accuracy Score is 0.9464285714285714
```

Figure 11. CNN confusion matrix.

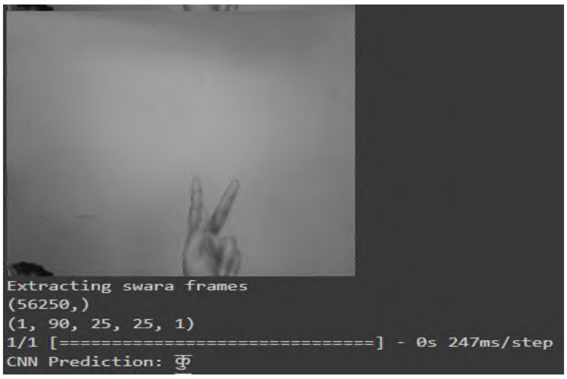


```
Extracting swara frames
(56250,)
(1, 90, 25, 25, 1)
1/1 [==============================] - 0s 477ms/step
CNN Prediction: का
```

Figure 12. 3D CNN prediction for letter “Ka”.

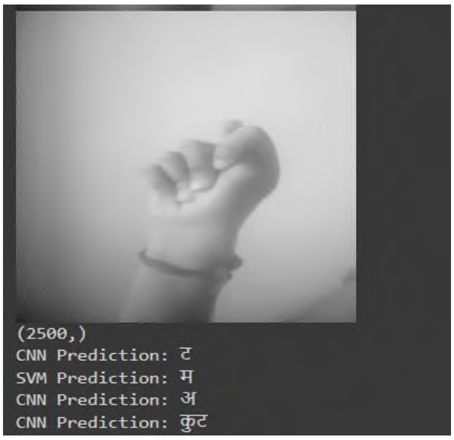
Table 3. Classification report of 3D CNN.

	Precision	Recall	F1-score	Support
0	1.00	0.80	0.89	10
1	1.00	1.00	1.00	6
2	0.60	1.00	0.75	3
3	1.00	1.00	1.00	2
4	1.00	1.00	1.00	6
5	1.00	1.00	1.00	3
-	-	-	-	-
Accuracy	-	-	0.93	30
Macro average	0.93	0.97	0.94	30
Weighted average	0.96	0.93	0.94	30



```
Extracting swara frames
(56250,)
(1, 90, 25, 25, 1)
1/1 [==============================] - 0s 247ms/step
CNN Prediction: कु
```

Figure 13. 3D CNN prediction for letter “Ku”.



```
(2500,)
CNN Prediction: ट
SVM Prediction: म
CNN Prediction: अ
CNN Prediction: कुट
```

Figure 14. 3D CNN prediction for word “Kut”.

5. Conclusion

Sign language recognition using machine learning and deep learning techniques has great potential to improve the lives of people with hearing and speaking impairments. This process involves several important steps, including dataset creation, data augmentation, model building, video processing, image preprocessing, Vyanjan determination, feature extraction, and image classification. We can construct precise and dependable models that recognize sign language gestures with high accuracy by carefully performing each step. Video preprocessing for Swara determination, concatenating Vyanjan with Swara to determine letters, and using transliteration in python to display the output in Marathi are critical steps in accurately recognizing letters in a video. The use of advanced techniques such as background subtraction and 3D CNNs helps to improve the accuracy of the model. The use of transliteration helps to make the output more accessible and user-friendly. The recognized text is then used as a query to search engines to extract relevant search results based on the user’s requirements. In the future, the prediction system can be improved by increasing the veracity of training data.

6. Future scope

The proposed system consists of 6 output classes. In future, it can be expanded to 43 output classes for every Mulakshar and Swar in Marathi language.

Author contributions

Conceptualization, SM, KG, VG, SK and KS; methodology, SM, KG, VG, SK and KS; software, SM, KG, VG, SK and KS; validation, SM, KG, VG, SK and KS; formal analysis, SM, KG, VG, SK and KS; investigation, SM, KG, VG, SK and KS; resources, SM, KG, VG, SK and KS; data curation, SM, KG, VG, SK and KS; writing—original draft preparation, SM, KG, VG, SK and KS; writing—review and editing, SM, KG, VG, SK and KS; visualization, SM, KG, VG, SK and KS; supervision, SM, KG, VG, SK and KS; project administration, SM, KG, VG, SK and KS; funding acquisition, SM, KG, VG, SK and KS.

Conflict of interest

The authors declare no conflict of interest.

References

1. Katoch S, Singh V, Tiwary US. Indian sign language recognition system using SURF with SVM and CNN. *Array* 2022; 14: 100141. doi: 10.1016/j.array.2022.100141
2. Huang J, Zhou W, Li H, Li W. Sign language recognition using 3D convolutional neural networks. In: Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME); 29 June–3 July 2015; Turin, Italy.

3. Haldera A, Tayadeb A. Real-time vernacular sign language recognition using mediapipe and machine learning. *International Journal of Research Publication and Reviews* 2021; 2(5): 9–17.
4. Mali D, Limkar N, Mali S. Indian sign language recognition using SVM classifier. In: Proceedings of the International Conference on Communication and Information Processing (ICCIP); 18 May 2019.
5. Patil R, Patil1 V, Bahuguna A, Datkhile G. Indian sign language recognition using convolutional neural network. In: *ITM Web of Conferences*, Proceedings of the International Conference on Automation, Computing and Communication 2021 (ICACC-2021); 14–15 July 2021; Nerul, India. EDP Sciences; 2021. Volume 40.
6. Rani RS, Rumana R, Prema R. A review paper on sign language recognition for the deaf and dumb. *International Journal of Engineering Research & Technology (IJERT)* 2021; 10(10). doi: 10.17577/IJERTV10IS100129
7. Sood D. Sign language recognition using deep learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* 2022; 10(Ⅲ). doi: 10.22214/ijraset.2022.40627
8. Darekar AA, Pawar NB, Pawar RD, et al. Marathi sign language recognition for physically disabled people. *International Journal of Advanced Research in Science, Communication and Technology* 2022; 2(7). doi: 10.48175/IJARSCT-4435
9. Shinde A, Kagalkar RM. Advanced Marathi sign language recognition using computer vision. *International Journal of Computer Applications* 2015; 118(13): 1–7. doi: 10.5120/20802-3485
10. Subramanian B, Olimov B, Naik SM, et al. An integrated mediapipe-optimized GRU model for Indian sign language recognition. *Scientific Reports* 2022; 12(1): 11964. doi: 10.1038/s41598-022-15998-7
11. Daroya R, Peralta D, Naval P. Alphabet sign language image classification using deep learning. In: Proceedings of the TENCON 2018–2018 IEEE Region 10 Conference; 28–31 October 2018; Jeju, Korea.
12. Patravali1 SD, Wayakule JM, Katre AD. Skin segmentation using YCBCR and RGB color models. *International Journal of Advanced Research in Computer Science and Software Engineering* 2014; 4(7): 341–346.
13. Mohamed SS, Tahir NM, Adnan R. Background modelling and background subtraction performance for object detection. In: Proceedings of the 2010 6th International Colloquium on Signal Processing & Its Applications; 21–23 May 2010; Malacca, Malaysia.