Heliyon 10 (2024) e38265

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Derived Amharic alphabet sign language recognition using machine learning methods

Ayodeji Olalekan Salau [a,e,*], Nigus Kefyalew Tamiru [b], Bekalu Tadele Abeje [c,d]

[a] Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria
[b] Department of Electrical and Computer Engineering, Debre Markos University, Debre Markos, Ethiopia
[c] Bahir Dar Institute of Technology Department of Computer Science, Bahir Dar, Amhara, Ethiopia
[d] Department of Information Technology, Haramaya University, Dire Dawa, Ethiopia
[e] Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India

ARTICLE INFO

Keywords:
Derived Amharic alphabet signs
CNN
HOG
SVM

ABSTRACT

Hearing-impaired people use sign language as a means of communication with those with no hearing disability. It is therefore difficult to communicate with hearing impaired people without the expertise of a signer or knowledge of sign language. As a result, technologies that understands sign language are required to bridge the communication gap between those that have hearing impairments and those that dont. Ethiopian Amharic alphabets sign language (EAMASL) is different from other countries sign languages because Amharic Language is spoken in Ethiopia and has a number of complex alphabets. Presently in Ethiopia, just a few studies on AMASL have been conducted. Previous works, on the other hand, only worked on basic and a few derived Amharic alphabet signs. To solve this challenge, in this paper, we propose Machine Learning techniques such as Support Vector Machine (SVM) with Convolutional Neural Network (CNN), Histogram of Oriented Gradients (HOG), and their hybrid features to recognize the remaining derived Amharic alphabet signs. Because CNN is good for rotation and translation of signs, and HOG works well for low quality data under strong illumination variation and a small quantity of training data, the two have been combined for feature extraction. CNN (Softmax) was utilized as a classifier for normalized hybrid features in addition to SVM. SVM model using CNN, HOG, normalized, and non-normalized hybrid feature vectors achieved an accuracy of 89.02%, 95.42%, 97.40%, and 93.61% using 10-fold cross validation, respectively. With the normalized hybrid features, the other classifier, CNN (sofmax), produced a 93.55% accuracy.

1. Introduction

There are numerous sign languages in use around the world, including American sign language, Indian sign language, British sign language, French sign language, Chinese sign language, and Ethiopian sign language (ESL). Since sign language (SL) is not a universal language, each country has its own SL and sign language alphabets [1–3]. Sign language has its own syntactical and grammatical meaning, which differs from country to country. Researchers have utilized a variety of approaches to recognize sign languages and hand gestures. Isolated or continuous signs, numbers or alphabets can all be used in sign language [4]. Single sign is presented in an isolated sign system, whereas continuous sign recognition, which is a complete or full sentence is supplied in the form of continuous

* Corresponding author. Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria.
E-mail addresses: ayodejisalau98@gmail.com (A.O. Salau), kefyalewnigus@gmail.com (N.K. Tamiru), bekalutadele@gmail.com (B.T. Abeje).

https://doi.org/10.1016/j.heliyon.2024.e38265
Received 1 May 2024; Received in revised form 19 September 2024; Accepted 20 September 2024
Available online 21 September 2024

List of abbreviations

ANN	Artificial Neural Network
AMASLR	Amharic Alphabet Sign Language Recognition
ASL:	American Sign Language
CNN	Convolutional Neural Network
DWT	Discrete Wavelet Transform
ESL:	Ethiopian Sign Language
ESLR	Ethiopian Sign Language Recognition
GF	Gabor Filter
HCI	Human–Computer Interaction
HOG	Histogram of Oriented Gradients
LBP	Local Binary Pattern
PCA	Principal Component Analysis
SLR	Sign Language Recognition
SVM	Support Vector Machine

signs (a static or moving number or alphabet sign is possible). Only the letters 'J' and 'Z' in American Sign Language (a-z) require motion, whereas the remaining 24 letters are presented as static images [5]. In Ethiopia, the first (Gee'z) 34 basic Amharic alphabet (Fidel) signs are static signs, while all of the derived alphabet signs are dynamic signs.

According to Refs. [6–8], around 466 million people worldwide suffer from hearing loss. Children make up 34 million of these people. World Health Organization has estimated that over 900 million people will suffer from hearing impairment by 2050. As a results of this, by 2050, the communication gap is most likely to increase in countries with limited combative technologies. As a result, technologies that recognizes sign language are required to bridge this communication gap. Many studies on Sign Language Recognition (SLR) have been conducted using vision-based and sensor-based methods. Nowadays, vision based SLR is a highly active area of research. However, Amharic Alphabet Sign Language Recognition (AMASLR) has received little attention. As a result of this, conducting research on AMASLR is a viable option.

The rest of this paper is organized as follows. Section 2 presents a review of related works, while section 3 presents about Ethiopian Sign Language based on bastard Amharic alphabet signs. The proposed system model is described in section 4. Section 5 presents about experimentation and result discussion and finally, section 6 concludes the conclusion and future work parts.

2. Related works

Sign Language Recognition (SLR) contributes to enhancing the development of human–computer interaction (HCI) systems. Many studies have been done on Sign Language Recognition utilizing machine learning approaches. In this section, we give a brief discussion of related works.

Recognition of Amharic sign language with Amharic alphabet signs using ANN and SVM was proposed in Ref. [2]. The paper presented the development of an automatic Amharic sign language translator which translates Amharic alphabet signs into their corresponding text using digital image processing and machine learning algorithms. The proposed system has four major developmental stages which include preprocessing, segmentation, feature extraction and classification. A total number of thirty-four features were extracted from shape, motion and color of hand gestures to represent both the base and some derived class of Amharic sign characters. Classification models were built using artificial neural network (ANN) and multi-class support vector machine (SVM). The results show that the recognition system is capable of recognizing the Amharic alphabet signs with an average accuracy of 80.82 % and 98.06 % using the ANN and SVM classifiers, respectively.

In [9], the authors used SLR to recognize Indian sign language from video clips. To extract features, the authors combined Discrete Wavelet Transform (DWT) with Local Binary Pattern (LBP) and employed ANN for classification and achieved a 92.79 % accuracy of recognition. In addition, a classification experiment was carried out using SVM, which yielded better results. In Ref. [10], a Persian Sign Language trajectory-based recognition system which comprised of 1200 videos from 12 signers was developed for 20 dynamic signs. The features that were used to train the model were the frame's centroid, area, eccentricity, and orientation. Finally, HMM classifier was used which achieved an average recognition rate of 98.13 %, whereas the SVM classifier achieved an average recognition rate of 87.77 %. HOG based Single-Handed Bengali SLR was used for 35 Bengali alphabets. Each class contains 40 images. Resizing, Histogram equalization, smoothing, skin segmentation and color conversation are included in the preprocessing stage. KNN classifier with HOG feature extractor achieved 91.1 % accuracy. Authors in Ref. [11] proposed a SLR method which used deep learning to recognize static alphabets of American Sign Language (ASL). For feature extraction and classification, Convolutional Neural Network (CNN) was used. The experiment was conducted on static alphabets, numbers and static words. The authors obtained 99 % and 90.04 % training and testing accuracy respectively. The authors in Ref. [12] presented an Indian sign language gesture recognition system based on deep learning and image processing to recognize static, dynamic, and number signs. For 36 static gestures, 45000 RGB images were gathered, as well as 1080 videos for 10 dynamic gesture words. To train the videos of dynamic gestures to obtain the static gestures, the authors employed CNN with Softmax Classifier. The training accuracy for the 36 static gesture signs was 98.81 %, and the training accuracy for the 10 dynamic gesture signs was 99.08 %.

Only a few studies have been conducted in Ethiopia to recognize Amharic alphabet sign language using a vision-based approach. Zerubbabel [13], developed an AMSL recognition system for only 10 fundamental alphabets that are static, easy to sign and convert to text. Some of the processing stages performed include Color conversion from RGB to grayscale, contrast control, and sharpening. The system achieved a recognition rate of 88.08 % using neural network with PCA driven features and 96.22 % recognition rate using neural network with haar-like features. Admasu and Raimond [14] presented a hand gesture recognition system that used Gabor Filter (GF) with Principal Component Analysis (PCA) for feature extraction and ANN for recognition of Ethiopian Sign Language (ESL) of the basic 34 letters of the Ethiopian Manual Alphabet (EMA), with a 98.53 %. They did not, however, take into account any derived letters (Fidels). In Ref. [15], the authors presented an offline candidate hand gesture selection and trajectory determination for continuous ESL. The focus of this research was on a system that extracts candidate EMA frames from a video sequence and calculates hand movement trajectories. The designed system has two separate parts; the Candidate Gesture Selection (CGS) and the Hand Movement Trajectory Determination (HMTD). The CGS combines speed profile of continuous gestures and Modified Housdorff Distance (MHD) measure to achieve a 80.72 % accuracy. The HMTD was performed by considering each hand gesture centroid from frame to frame and using angle, x- and y-directions, with an accuracy of 88.31 %. The overall system performance is 71.88 %.

2.1. *Amharic alphabet signs*

There are different basic and derived Amharic alphabet signs. The basic alphabet signs have their own first, second, third, fourth, fifth, sixth, and seventh derived alphabet signs. The basic signs are represented by single static hand shape images, while the derived signs are dynamic signs which need hand shape movement [2,13,14]. Some basic and their own derived Amharic alphabet are presented in Table 1.

The Ethiopian Sign Language (ESL) was created effectively to help hearing-impaired people in the society. Thus, Amharic Sign Language recognition is used in Ethiopia to help hearing-impaired people to effectively communicate. In this study, we only included seven derived Amharic alphabet signs. Fig. 1 shows the seven Ethiopian derived Amharic alphabet finger spellings.

In Fig. 1, the derived Amharic alphabet signs of ESL are presented. These signs are ሏ (lua), ሟ (mua), ሯ (rua), ሷ (sua), ሿ (shua), ቋ (kua), and ቧ (bua). They are dynamic single hand shapes and the remaining signs also make use of dynamic single hand sign notations.

3. The proposed system architecture

The three major stages of the proposed system architecture include image processing (preprocessing, segmentation, and feature extraction), model construction, and testing. The image processing techniques have the capability of enhancing the different images using preprocessing, segmentation, and feature extraction activities for further process. Using feature vectors, the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models are developed after several training. Finally, both models (SVM and CNN) are tested. The proposed system architecture is shown in Fig. 2.

3.1. *Video acquisition*

For this study, there is no standard dataset/corpus of derived Amharic Alphabet Signs in Ethiopian Sign Language. Therefore, the dataset was acquired from voluntary signers whom comprise of experts and non-experts using different Smart Mobile Phone to make a robust database.

3.2. *Image processing*

There are three types of image processing techniques which were used in this study. They include image preprocessing, image segmentation, and feature extraction. Image preprocessing includes video to frame conversion, key frame selection, concatenation of key frames, and resizing the acquired images. Segmentation was achieved by RGB to grayscale conversation, image enhancement, noise removal, and sharpening methods. CNN, HOG features, Hybrid feature (HOG and CNN) extraction, and feature selection using ANOVA were performed in the feature extraction stage to create a knowledge base or feature set. After several training, SVM and CNN models are constructed and afterwards tested by unknown sample data.

Table 1
Sample basic and derived Amharic alphabets.

Basic Amharic alphabet	1st derived alphabet	2nd derived alphabet	3rd derived alphabet	4th derived alphabet	5th derived alphabet	6th derived alphabet	7th derived alphabet
ለ (le)	ሉ (lu)	ሊ (li)	ላ (la)	ሌ (lie)	ል (l)	ሎ (lo)	ሏ (lua)
መ (me)	ሙ (mu)	ሚ (mi)	ማ (ma)	ሜ (mie)	ም (m)	ሞ (mo)	ሟ (mua)
ረ (re)	ሩ (ru)	ሪ (ri)	ራ (ra)	ሬ (rie)	ር (r)	ሮ (ro)	ሯ (rua)
ሰ (se)	ሱ (su)	ሲ (si)	ሳ (sa)	ሴ (sie)	ስ (s)	ሶ (so)	ሷ (sua)
ሸ (she)	ሹ (shu)	ሺ (shi)	ሻ (sha)	ሼ (shie)	ሽ (sh)	ሾ (sho)	ሿ (shua)
ቀ (ke)	ቁ (ku)	ቂ (ki)	ቃ (ka)	ቄ (kie)	ቅ (k)	ቆ (ko)	ቋ (kua)
በ (be)	ቡ (bu)	ቢ (bi)	ባ (ba)	ቤ (bie)	ብ (b)	ቦ (bo)	ቧ (bua)

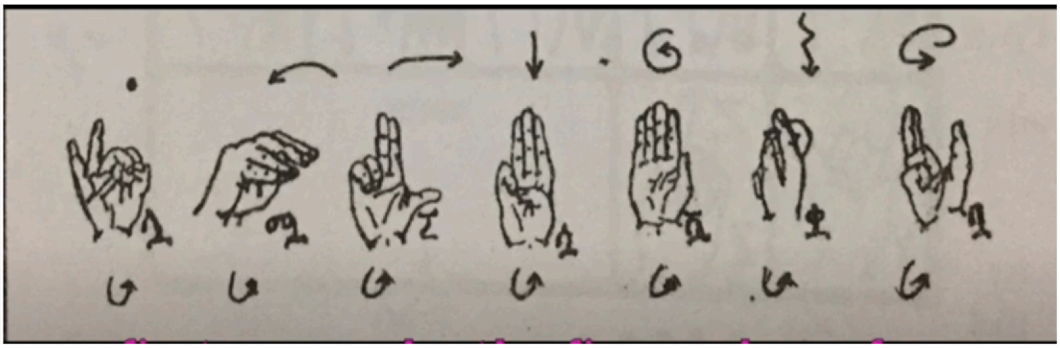

Fig. 1. The seventh derived sample Amharic figure spelling.

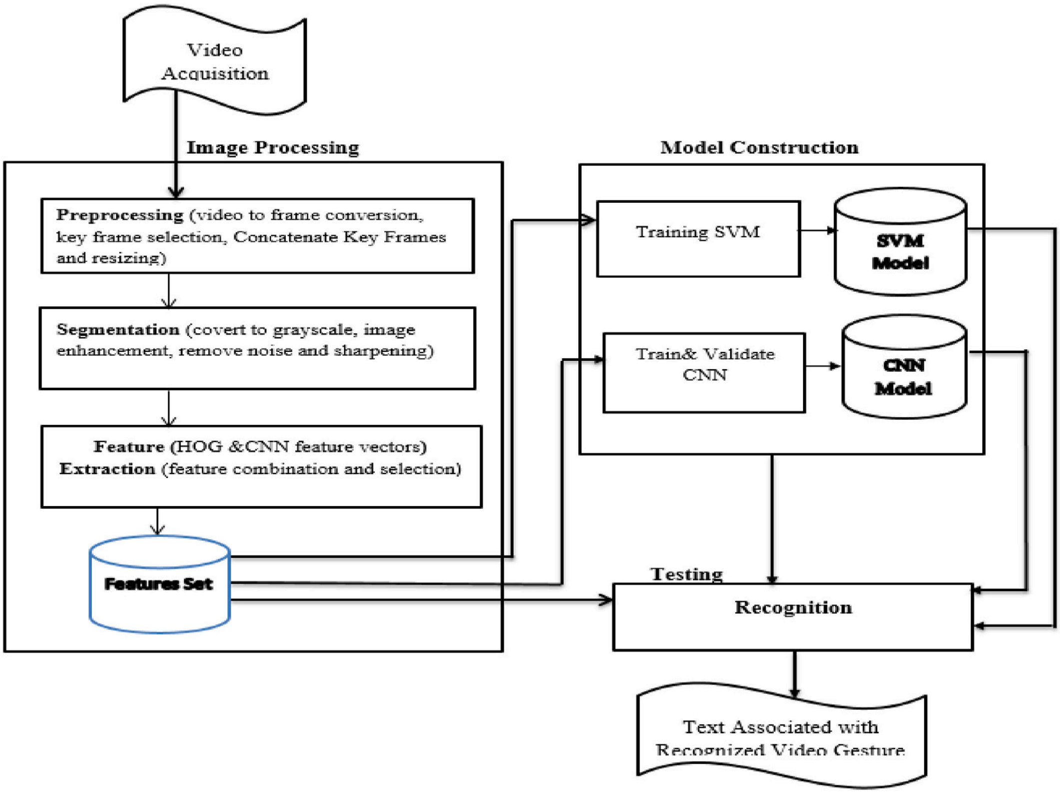

Fig. 2. The proposed system architecture.

Table 2
Video to frame conversion algorithm (Algorithm 1).

Algorithm for Video to frame conversion
Input: *captured video*
Output: *the sequence of frames*
Begin:
Input video
Convert to frame by frame
Return the sequence of frames
End:

3.2.1. Preprocessing

A Video to Frame Conversion

After we acquired the seven derived Amharic Alphabet Signs, the acquired videos were converted into several frames using frame by frame conversion methods. The recorder records at 30 frames per second, so a second video was used to create thirty frames or images. The algorithm for video to frame conversion is presented in Table 2.

B Key Frame Selection

Key frames are the desired number of frames that have significance impact on the video frames. The authors in Ref. [16] proposed structural similarity to calculate the similarity between two images which can be used as a key frame selection and redundant frame removing technique. However, this approach doesn't work for Amharic alphabet sign language because there is no structural difference between the first basic and the derived alphabet signs; thus, the mean square error method was employed. Therefore, we used the key frame selection method which is also used as a down sampling technique in order not to lose a significant number of frames. Thus, the Mean square algorithm was performed on the mean square error of the pixel difference between consecutive frames and then followed by bubble sort strategy for down sampling. The algorithm is presented in Table 3.

By using the Mean squared error (MSE), we obtain the difference between each pixel value from the predecessor and the successor frame as given by Eq. (1) [17]. It is used to calculate the difference in value between each pixel intensity values from one frame to another frame in as much that they have the same shape. Lower mean square error indicates the similarity, while higher mean square error indicates a large differences of the image set. Furthermore, similar images have a mean square error of zero.

$$\boldsymbol{MSE}=\frac{\mathbf{1}}{\boldsymbol{mn}}\sum_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{m}-\mathbf{1}}\sum_{\boldsymbol{j}=\mathbf{0}}^{\boldsymbol{n}-\mathbf{1}}[\boldsymbol{I}(\boldsymbol{i},\boldsymbol{j})-\boldsymbol{K}(\boldsymbol{i},\boldsymbol{j})]^{\mathbf{2}} \tag{1}$$

C Frame Concatenation

Concatenation in this case simply represents conversion of the videos into a single video with a single image patterns. Frames can be concatenated vertically, horizontally, and by using square method. For this work, horizontal concatenation was implemented because it is more visible even with the naked eye to represent the motion of the derived Amharic alphabet signs. Sample frames are extracted and then concatenated starting from the letter 'ሯ''s video as shown in Fig. 3.

D Image resizing

The dataset contains images with various sizes. Therefore, the images need to be reduced to make their size more appropriate for further process. The dynamic derived Amharic alphabet signs were concatenated and then reduced to smaller sizes which helps to reduce computational complexity and processing time.

3.2.2. Segmentation

Segmentation refers to the separation of significance foreground image from the background object. Color conversion, image enhancement, noise filtering, and sharpening activities in the segmentation stage are described in (A) to (C).

A ***Color Conversion:*** Gray scale images require low computational time than RGB image. Due to this, the RGB images were converted to grayscale by using cv2.COLOR_BGR2GRAY python function.
B ***Image Enhancement:*** After the RGB color frames/images are converted to grayscale frames, image/frame enhancement is required to improve the quality and adjust the contrast of the grayscale images in order not to affect the performance of the Amharic SLR system. The quality of the images are seen to improve and become easier to process.
C ***Filtering Noise and Sharpening:*** Filtering of noise is used to enhance the data and it allows accurate representation of pixels. Noise removal algorithms remove or reduce the visibility of noise by smoothing the entire image leaving areas near contrast boundaries.

Table 3
Mean square algorithm (Algorithm 2).

Algorithm for finding Mean Square Errors
Input: *The sequence of frame Difference*
Output: *Error*
def MSE (frame1, frame 2):
*squared_difference = (frame1 - frame2) ** 2*
sum = sum(squared_difference)
*number_of_pixel = frame1.shape[0] * frame1.shape [1]*
error = sum/number_of_pixel
return error

Fig. 3. Sample of concatenated image.

For this work, median filter is used to remove noise because it allows image edge in efficient performance for low density of noise without affecting the sharpness of the image. For noise removal, **cv2.medianBlur (image, kernel size)** function with a kernel size of three was performed.

On the other hand, sharpening is an effect added to images to make the images to have a sharper appearance or make edges visible because it increases the sharpeness and local contrast of the image. Human vision is sensitive to fine details and edges of an image, and the pictorial quality of an image can be extremely corrupted if high frequencies are mitigated or completely removed.

3.2.3. Feature extraction

There is no generic feature extraction scheme that works in all cases because it depends on the type of problem. For this work, CNN, HOG, normalized and non-normalized hybrid features (CNN and HOG features) are implemented. CNN features extraction technique is used to extract deep features. However, hand crafted features are extracted by HOG. Therefore, Hybrid feature extraction approach is the combination of hand crafted and deep features. And the other, the study used ANOVA which is feature selection algorithm to select significance features. Hybrid feature extraction and feature selection techniques are briefly explained in detail bellow.

A Hybrid feature extraction

Hybrid feature vector is a combination of two or more feature vectors that are extracted with different feature extractor algorithm, in our case, deep features with CNN and Handcrafted features with HOG.

Nowadays, Deep feature extraction has dominated Hand crafted feature extraction. However, deep feature extraction using CNN is not good when the image has high illumination variation and low quality. Therefore, it is the view of the authors that handcrafted feature extraction is better than deep feature extraction, especially because HOG is known to be highly resistant to illumination variation because of its block normalization and because it has its own gamma correction and histogram equalization function. Therefore, we proposed hybrid feature extraction to overcome the weakness of individual feature extraction approaches. However, CNN is good for translation as well as rotation variation and HOG is good for illumination variation and low quality images [18].

B Feature selection

ANOVA feature selection algorithm has been used for feature selection because there are different feature vector lengths in this prototype. Therefore what we done was to reduce the longest feature vector and make equal the shortest feature vector. ANOVA is also a robust technique; it assumes all sample of data to be distributed in general, having equal variance and independence. In the feature selection process, select K Best class with **score_func=f_classif statical test** of the python function which is used to select different significance features.

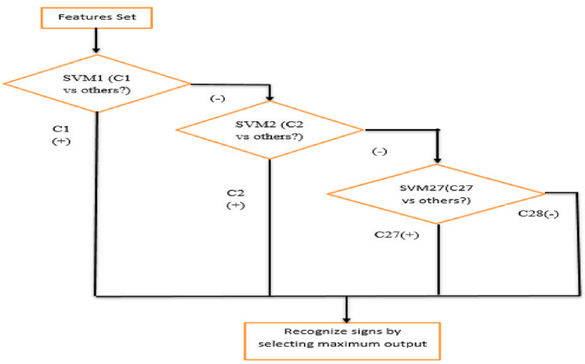

Fig. 4. One-versus-rest multiclass SVM structure for classification.

3.2.4. Model construction

A Support Vector Machine Modeling

After feature vectors are extracted using HOG, CNN, normalized and non-normalized hybrid features (CNN and HOG features), multi class SVM is used for classification. SVM gives more satisfactory result in image classification and recognition.

The performance of CNN softmax and SVM comparison is presented in Ref. [19] for Non-Touch sign word recognition and for Hyperspectral Image Classification. The result shows that the SVM classifier outperforms the CNN softmax classifier. In N-dimensional (number of features) space, SVM approach builds a hyperplane for classification and regression. This was done using a hyperplane, which was the largest categorization for every category's nearby training data points. In this research, we utilized a multiclass SVM that employed labels from the feature vector to classify the data since the study has several number of classes. So, we selected one versus the rest (OVR) SVM technique which assigned the classified class with the highest output function. In this scenario, the kernel function was used to categorize non-linear datasets and convert them to linear datasets. For the SVM's functionality, we applied the radial basis kernel (RBF) that may provide localization and restricted response across the entire range of the main axis. As a result, multi-class OVR SVMs ran in parallel, which is demonstrated in Eqs. (2) and (3), one class was isolated from the others. This is done using each of the support vectors. Fig. 4 shows the multiclass SVM approach.

$$f_{i(x)=w_i^t x+b_i} \tag{2}$$

$$X\rightarrow arg\ max_{if_i\ \ (x)} \tag{3}$$

where the ith decision function classifies class i with positive labels and remaining classes as negative labels. $fi(x)$ is the N-dimensional vector and bi is scalar, and X classifies the maximum output.

B Conventional Neural Network (CNN)

CNN is a type of deep learning algorithm used for both feature extraction and classification. The last layers of CNN is fully connected layer (dense layer) that is used for classification purpose. The fully connected layer is used to compute the final output probabilities for each class. Fully connected layer applied at the end of CNN model before applying the classifier layer usually soft max which is the last layers of CNN for classification. The three fully connected layers in our work and the output of their layer is the input of the softmax classifier.

For this study, we used one of adaptive learning rate estimation methods because it is not affected by the type of model and problem as indicated in Refs. [20–22]. The main target of the optimization algorithm (Adam) is reducing the difference between actual output and predicated output that is cost function or loss function and it computes the network loss function by calculating the estimation of individual adaptive learning rate from the parameters of the first and second moments of the gradient.

3.2.5. Testing

The recognition and classification accuracy of both built models, CNN and SVM, were assessed in this stage. The k-fold cross-validation approach (k = 10) is a good way to test the performance of CNN and SVM models [2]. Consequently, we choose k-fold cross-validation technique because all datasets were used for both training and validation. The entire dataset is divided into ten equal parts. Each partition served as both a training and an evaluation tool [23–25]. The total used dataset in this study is 2430, as described in Section 5. It was extracted in 1764 feature vectors being fed into the training machines.

4. Results and discussion

This section presents the discussion on the acquired dataset, the performance evaluation of SVM with CNN, HOG, normalized and non-normalized features, and also about CNN classifier with normalized hybrid features on k-fold (k = 10) cross-validation. This is an extensive experimental evaluation of the proposed method.

4.1. Dataset

When we were recording the dataset, fifteen voluntary signers were involved after giving 15 min training on display board. Ten of them were experts (five males and five females, from Debre Markos teaching college) and the rest persons were not experts. For the seventh derived Amharic alphabets, each signer signs 6 times, in total 90 data per each sign. Finally, we prepared 2430 dataset.

Table 4
10-Fold cross validation of SVM method using CNN and HOG features.

Fold	1	2	3	4	5	6	7	8	9	10	mean
Accuracy (SVM With CNN Features) %	90.09	88.32	88.92	89.62	89.27	87.62	90.33	88.92	87.26	88.96	**89.02**
Accuracy (SVM With HOG Features) %	95.19	95.85	93.88	95.75	94.34	96.79	96.04	95.57	95.75	95.09	95.42

4.2. Performance evaluation

4.2.1. SVM model with CNN features

For SVM classifier, the basic parameters are Kernel, C (penalty parameter), gamma value and decision making function shape. By using grid search algorithm kernel = 'rbf', C = 100, gamma='auto', decision_function_shape = 'ovr', and cache_size = 300 gives an optimal value. Table 4 shows that the accuracy result of SVM for each fold.

4.2.2. SVM model with HOG features

The SVM classifier was trained and tested with HOG features by feeding 1764 features after feature selection. Each performance result of the SVM model with HOG features on each fold is presented in Table 4. As observed from Table 4, the minimum accuracy was recorded on fold 9 (87.26 %) and the maximum accuracy was obtained on fold 1 (90.09 %) with a 2.83 % difference. The mean accuracy of the SVM classifier with CNN features using 10-fold is 89.02 % which is a good result. On the other hand, the SVM model with HOG features achieved 93.88 % minimum performance on fold 3 and 96.79 % maximum performance on fold 6. A 2.58 % difference was achieved between them. This shows that, the mean accuracy of the SVM model with HOG features on 10-fold cross validation is 95.42 %, which is better than the mean accuracy of the SVM model with CNN features whose performance is 89.02 %. In this work, we used different measurement techniques such as Precision, recall, F1-Score, and average accuracy. The SVM classifier results with CNN and HOG features are presented in Table 5.

4.2.3. SVM model with non-normalized hybrid (CNN and HOG) features

The SVM model was evaluated on non-normalized hybrid features. The length of hybrid features, CNN and HOG after feature selection was reduced to small size feature vectors and then those reduced features were feed to the SVM classifier. The performance was measured by the mean of the accuracy that was obtained by the 10-fold cross validation.

4.2.4. SVM model with normalized hybrid features

The main aim of hybrid feature normalization is to make the features comparable with each other, which gives an opportunity for all features to participate for the final decision or classification. We used min-max feature normalization because it preserves the relationships among the original data values. Table 6 shows that the SVM model was trained and tested with normalized and non-normalized hybrid feature vectors on 10-fold cross validation.

Table 6 presents the SVM model with non-normalized hybrid features of minimum accuracy on fold 6 and the maximum accuracy

Table 5
Results of SVM with CNN and HOG features.

SVM with CNN features				**SVM with HOG features**		
7th derived alphabet signs	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**
ሏ	0.98	0.85	0.85	0.98	1.00	0.99
ሗ	0.97	0.95	0.96	0.97	0.95	0.96
ሟ	0.99	0.97	0.96	0.95	0.97	0.96
ሧ	0.98	0.75	0.75	0.98	1.00	0.99
ሯ	0.97	0.97	0.97	0.97	0.97	0.97
ሷ	1	0.8	0.75	1.00	1.00	1.00
ሿ	1	0.95	0.97	1.00	0.95	0.97
ቋ	0.74	0.76	0.84	1.00	1.00	1.00
ቧ	0.95	0.97	0.96	0.95	0.97	0.96
ቯ	0.97	0.95	0.96	0.97	0.95	0.96
ቷ	0.78	0.82	0.91	0.90	0.93	0.91
ቿ	0.86	0.93	0.89	0.86	0.93	0.89
ኋ	0.93	0.8	0.8	0.93	1.00	0.96
ኗ	0.95	0.97	0.96	0.95	0.97	0.96
ኟ	0.99	0.97	0.99	0.91	0.97	0.94
ዟ	0.75	0.82	0.74	1.00	1.00	1.00
ዧ	0.98	0.95	0.88	0.98	1.00	0.99
ዷ	0.97	0.95	0.96	0.97	0.95	0.96
ጇ	0.85	0.88	0.76	1.00	1.00	1.00
ጓ	0.97	0.97	0.95	0.97	0.97	0.97
ጧ	0.85	0.84	0.78	1.00	1.00	1.00
ጯ	0.95	0.98	0.98	0.95	1.00	0.98
ጿ	0.8	0.9	0.98	1.00	0.90	0.95
ኧ	0.85	0.88	0.91	0.95	0.88	0.91
ፏ	0.95	0.88	0.98	0.95	1.00	0.98
ፗ	0.8.	0.95	0.97	1.00	0.95	0.97
ኳ	0.75	0.95	0.98	0.90	0.95	0.93
accuracy			**0.90**			**0.962**
macro avg	**0.91**	**0.90**	**0.90**	**0.962**	**0.962**	**0.962**
weighted avg	**0.91**	**0.90**	**0.90**	**0.962**	**0.962**	**0.962**

Table 6
10-Fold cross validation of SVM with normalized and non-normalized hybrid features.

Fold	1	2	3	4	5	6	7	8	9	10	Mean
Accuracy (SVM with non-normalized hybrid features)	93.30	91.79	91.04	92.83	92.36	95.57	95.75	95.85	94.06	93.58	**93.61**
Accuracy (normalized hybrid features)	96.79	96.60	96.51	98.02	96.51	98.58	98.49	97.17	97.45	97.83	97.40

on 3-fold are 95.57 % and 91.04 % respectively. Their fold accuracy difference is 4.81 %. The mean accuracy of this model is 93.61 %. However, the minimum accuracy of the same model with different features which are normalized hybrid features on both fold 3 and fold 5 is 96.51 %, whereas the maximum accuracy, 98.58 % was achieved on fold 6. The difference accuracy value of fold 3 or 5 and 6 is 2.84 % and the average performance, 97.40 % is obtained using this model. Thus, the mean accuracy of the SVM model with normalized features is better than the SVM with non-normalized features with 10-fold cross validation. The precision, recall and F-measure of SVM classifier with normalized and non-normalized hybrid feature vectors are presented in Table 7.

4.2.5. LeNet model and CNN classifier with normalized hybrid features

LeNet is a simple CNN architecture which is a type of feed-forward neural network. It contains two convolution layers, two average pooling layers, two fully connected layers and a softmax layer. It is robust to simple geometric transformations and distortions. The results of the experiments show that the accuracy of the LeNet model is 83.33 %, while the CNN model achieved 93.55 % accuracy which is better than the LeNet model on the normalized hybrid feature vectors as presented in Table 8.

4.3. Test results

This section illustrates the comparison results of SVM models on different features and determines which combination achieved a better accuracy. As described in section 4.2.5, the accuracy results of the SVM model with HOG, CNN, Normalized and non-normalized features (HOG and CNN feature vectors) are 95.42 %, 89.02 %, 97.40 % and 93.61 % respectively, while a 93.55 % performance result was obtained using the CNN (softmax) model with normalized features. Based on the normalized hybrid features, SVM proved to be a better classifier than CNN. The experiment results of the two models, SVM and CNN (sofmax) on several feature vectors is shown in Fig. 5.

Table 7
SVM with non-normalized and normalized hybrid feature vectors.

SVM with non-normalized hybrid features				**SVM with normalized hybrid features**		
7th **derived alphabet signs**	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**
ሏ	0.98	0.9	0.99	0.98	1.00	0.99
ሗ	0.97	0.95	0.96	0.97	0.95	0.96
ሟ	0.95	0.97	0.96	0.99	0.97	0.96
ሧ	0.98	0.95	0.99	0.98	1.00	0.99
ሯ	0.97	0.97	0.97	0.97	0.97	0.97
ሷ	0.97	0.97	0.81	1.00	1.00	1.00
ሿ	0.9	0.95	0.97	1.00	0.95	0.97
ቋ	0.9	0.9	0.8	1.00	1.00	1.00
ቧ	0.95	0.97	0.96	0.95	0.97	0.96
ቯ	0.97	0.95	0.96	0.97	0.95	0.96
ቷ	1	0.93	0.91	1.00	1.00	0.91
ቿ	0.76	0.93	0.89	0.86	0.93	0.89
ኋ	0.93	0.9	0.96	0.93	1.00	0.96
ኗ	0.85	0.97	0.96	0.95	0.97	0.96
ኟ	0.91	0.97	0.8	0.99	0.97	0.99
ዟ	0.91	0.91	0.91	1.00	1.00	1.00
ዧ	0.98	1	0.99	0.98	1.00	0.99
ዷ	0.97	0.95	0.96	0.97	0.95	0.96
ጇ	0.9	0.9	0.8	1.00	1.00	1.00
ጓ	0.97	0.97	0.97	0.97	0.97	0.97
ጧ	0.9	0.9	0.9	1.00	1.00	1.00
ጯ	0.95	0.9	0.98	0.95	1.00	0.98
ጿ	0.9	0.9	0.95	1.00	0.90	0.98
ፇ	0.95	0.88	0.91	0.95	0.88	0.91
ፏ	0.95	0.9	0.98	0.95	1.00	0.98
ፗ	0.9	0.85	0.96	1.00	0.95	0.97
ኳ	0.9	0.95	0.93	0.90	0.95	0.98
accuracy			**0.93**			**0.97**
macro avg	**0.930**	**0.92**	**0.93**	**0.97**	**0.97**	**0.97**
weighted avg	**0.930**	**0.93**	**0.93**	**0.97**	**0.97**	**0.97**

Table 8
LeNet model accuracy result and CNN classifier with normalized hybrid features (CNN and HOG).

Epoch No.	LeNet model		CNN model	
	Loss (%)	Accuracy (%)	Loss (%)	Accuracy (%)
1	0.2629	0.9169	0.122	0.9611
2	0.2999	0.9094	0.1652	0.9701
3	0.2617	0.9175	0.1178	0.986
4	0.5685	0.7611	0.1062	0.9881
5	0.6895	0.7454	0.4476	0.8838
6	0.6154	0.7687	0.3186	0.9235
7	0.37887	0.83226	0.2121	0.90245
8	0.31212	0.8812	0.3576	0.8748
9	0.4998	0.7901	0.2042	0.9151
10	0.3881	0.8101	0.1331	0.9501
Average	0.427679	0.833266	0.21844	0.935505

Note: Ten Samples of Epochs.

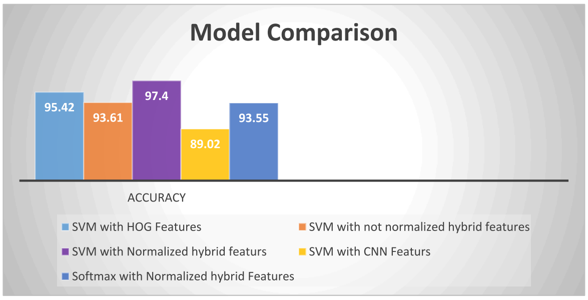

Fig. 5. SVM and CNN (softmax) models comparison.

4.4. Advancements beyond the state of the art

The current paper significantly achieved an advancement over the work of Nigus et al. [2]. While Nigus et al. [2] employed Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to achieve notable results in recognizing Amharic sign language, while this work introduced a novel hybrid feature extraction method that combines Convolutional Neural Networks (CNN) with Histogram of Oriented Gradients (HOG). This hybrid approach leverages on the strengths of both techniques, leading to superior accuracy and robustness in sign language recognition, especially under varying conditions.

4.4.1. Performance comparison

Nigus et al. [2] reported accuracies of 80.82 % using ANN and 98.06 % using SVM for the recognition of Amharic alphabet signs. In contrast, this research achieved higher performances, with SVM on normalized hybrid features achieving 97.40 % accuracy and CNN (softmax) achieving 93.55 %. These results indicate a significant improvement, particularly in recognizing derived Amharic alphabet signs, which are more complex and dynamic. The performance enhancement underscores the effectiveness of the hybrid feature extraction method and advanced classification models used in this research. A comparative analysis of the proposed method with existing works is presented in Table 9. The results show that the proposed method performed favorably as compared to other works.

5. Conclusion and future work

In this study, we developed a model for Amharic alphabet sign language recognition (AMASLR), which comprises of stages such as image processing techniques (such as skin segmentation, noise filtering, sharpening, and enhancing dataset images), feature extraction, and classification. For effective feature extraction, we combined CNN and HOG feature extraction methods. SVM and CNN (softmax) classifiers are used to recognize the extracted characteristics of the derived Amharic alphabet (ሆ, ኀ, ዶ, ቶ) signs, which produced good results. The experimental result demonstrates that the suggested model outperforms other models developed by earlier researchers. Even as we attained good results in recognizing AMASL based on the seven derived alphabet signs, it has its own impact on ESL. There are gaps which should be addressed in the future since the proposed system cannot be used as a full translation system for AMASL. Some of the recommendations for future work are as follows:

- Another challenge is the trajectory similarity among the dynamic signs, so we suggest that this problem may be solved if the researchers use a suitable trajectory determination algorithm.

Table 9
Comparative analysis of the proposed method with existing works.

Author	Method	Dataset	Accuracy	Remarks
[2]	Hybrid ANN and SVM	1710	80.82 % for SVM and 98 % for ANN	Implemented alphabet level for Amharic base and selected derived Amharic alphabets.
[14]	ANN	170	98.53 %	Implemented an image-based dataset for selected Amharic static alphabets.
[21]	End to End CNN	2550	98.5 %	Implemented only Amharic base alphabets with non-motional data.
Proposed	SVM and CNN	2430	SVM with HOG (95.42), SVM with CNN (93.55 %)	Implemented both static and derived sign language recognition of Amharic alphabets.

- The present study was done based on isolated derived alphabet sign, therefore, further work can be extended to continuous sign language recognition.
- The present study focused on seven derived Amharic alphabet signs. Therefore, we recommend to extend this work to word and sentence level.

Funding

Authors declare no funding for this research.

Availability of data

The datasets generated during and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Ayodeji Olalekan Salau: Writing – review & editing, Visualization, Methodology, Formal analysis, Conceptualization. **Nigus Kefyalew Tamiru:** Writing – original draft, Software, Resources, Methodology, Investigation, Data curation. **Bekalu Tadele Abeje:** Writing – original draft, Validation, Project administration, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

References

[1] K.K. Dutta, S. K, K. Raju, G.S. A. Kumar, S.A. Swamy B, Double handed Indian Sign Language to speech and text, 2015 Third International Conference on Image Information Processing (ICIIP) (2015) 374–377, https://doi.org/10.1109/ICIIP.2015.7414799.
[2] N.K. Tamiru, M. Tekeba, A.O. Salau, Recognition of Amharic sign language with Amharic alphabet signs using ANN and SVM, Vis. Comput. (2021) 1–16.
[3] S. Rathi, U. Gawande, Development of full duplex intelligent communication system for deaf and dumb people, in: 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, 2017, pp. 733–738.
[4] S. Fakhfakh, Y.B. Jemaa, Gesture recognition system for isolated word sign language based on key-point trajectory matrix, Comput. Sist. 22 (2018) 1415–1430.
[5] S. Shivashankara, S. Srinath, American sign language recognition system: an optimal approach, Int. J. Image Graph. Signal Process. 11 (2018) 18.
[6] S.I. Ertzgaard, N. Kristin, T. Sofie, H.G. Sindberg, H.T. Bang, M. Cosmas, et al., Prevalence of hearing impairment among primary school children in the Kilimanjaro region within Tanzania, Int. J. Pediatr. Otorhinolaryngol. 130 (2020) 109797.
[7] D. Maru, Deafness and hearing loss toolkit of educational resources, InnovAIT 14 (2021) 340–342.
[8] M. Spreckley, D. Macleod, B. González Trampe, A. Smith, H. Kuper, Impact of hearing aids on poverty, quality of life and mental health in Guatemala: results of a before and after study, Int. J. Environ. Res. Publ. Health 17 (2020) 3470.
[9] R. Sunitha, M. Suman, P.V.V. Kishore, K.K. Eepuri, Sign language recognition with multi feature fusion and ANN classifier, Turk. J. Electr. Eng. Comput. Sci. 26 (2018) 2871–2885.
[10] S.G. Azar, H. Seyedarabi, Trajectory-based recognition of dynamic Persian sign language using hidden Markov model, Comput. Speech Lang 61 (2020) 101053.
[11] T. Tabassum, I. Mahmud, M.P. Uddin, A. Emran, M.I. Afjal, A.M. Nitu, Enhancement of single-handed Bengali sign language recognition based on HOG features, J. Theor. Appl. Inf. Technol. 98 (2020) 743–756.
[12] N.K. Bhagat, Y. Vishnusai, G. Rathna, Indian sign language gesture recognition using image processing and deep learning, in: 2019 Digital Image Computing: Techniques and Applications (DICTA), 2019, pp. 1–8.
[13] L. Zerubabel, Ethiopian Finger Spelling Classification: a Study to Automate Ethiopian Sign Language, Addis Ababa University, Addis Ababa, Ethiopia, 2008. *Master's thesis*.
[14] Y.F. Admasu, K. Raimond, Ethiopian sign language recognition using Artificial Neural Network, in: 2010 10th International Conference on Intelligent Systems Design and Applications, 2010, pp. 995–1000.
[15] D. Abadi Tsegay, K. Raimond, Offline candidate hand gesture selection and trajectory determination for continuous Ethiopian Sign Language, J. Theor. Appl. Inf. Technol. 36 (2012).

[16] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[17] T. Meyer, Root mean square error compared to, and contrasted with, standard deviation, Survey. Land Inf. Sci. 72 (2012).

[18] M.I. Patel, V.K. Thakar, S.K. Shah, Image registration of satellite images with varying illumination level using HOG descriptor based SURF, Procedia Comput. Sci. 93 (2016) 382–388.

[19] M.A. Rahim, M.R. Islam, J. Shin, Non-touch sign word recognition based on dynamic hand gesture using hybrid segmentation and CNN feature fusion, Appl. Sci. 9 (2019) 3790.

[20] R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights into imaging 9 (2018) 611–629.

[21] B.T. Abeje, A.O. Salau, A.D. Mengistu, N.K. Tamiru, Ethiopian sign language recognition using deep convolutional neural network, Multimed. Tool. Appl. 81 (2022) 29027–29043, https://doi.org/10.1007/s11042-022-12768-5.

[22] A.O. Salau, N.K. Tamiru, D. Arun, Image-based number sign recognition for Ethiopian Sign Language using support vector machine, Lecture Notes in Electrical Engineering 925 (2022), https://doi.org/10.1007/978-981-19-4831-2_14. Springer, Singapore.

[23] A.A. Alemu, M.D. Melese, A.O. Salau, Ethio-Semitic language identification using convolutional neural networks with data augmentation, Multimed. Tool. Appl. 83 (2024) 34499–34514, https://doi.org/10.1007/s11042-023-17094-y.

[24] M.D. Melese, A.A. Alemu, A.O. Salau, I.G. Kasa, Speaker-based language identification for Ethio-Semitic languages using CRNN and hybrid features, Netw. Comput. Neural Syst. (2024) 1–23, https://doi.org/10.1080/0954898X.2024.2359610.

[25] A.A. Alemu, M.D. Melese, A.O. Salau, Towards audio-based identification of Ethio-Semitic languages using recurrent neural network, Sci. Rep. 13 (1) (2023) 19346, https://doi.org/10.1038/s41598-023-46646-3.