Hindawi
Advances in Human-Computer Interaction
Volume 2022, Article ID 1112169, 11 pages
https://doi.org/10.1155/2022/1112169

Research Article

A Generic Approach towards Amharic Sign Language Recognition

Netsanet Yigzaw,[1] **Million Meshesha**,[2] **and Chala Diriba**[1]

[1]*Faculty of Computing and Informatics, Department of Information Science (IKM), Jimma University, Jimma Institute of Technology, Jimma, Ethiopia*
[2]*School of Information Science, Department of Information Systems, Addis Ababa University, Addis Ababa, Ethiopia*

Correspondence should be addressed to Netsanet Yigzaw; missesyigzaw11@gmail.com

Received 19 May 2022; Accepted 15 July 2022; Published 22 September 2022

Academic Editor: Christos Troussas

In the day-to-day life of communities, good communication channels are crucial for mutual understanding. The hearing-impaired community uses sign language, which is a visual and gestural language. In terms of orientation and expression, it is separate from written and spoken languages. Despite the fact that sign language is an excellent platform for communication among hearing-impaired persons, it has created a communication barrier between hearing-impaired and non-disabled people. To address this issue, researchers have proposed sign language to text translation systems for English and other European languages as a solution. The goal of this research is to design and develop an Amharic digital text converter system using Ethiopian sign language. The proposed system was created with the help of two key deep learning algorithms: a pretrained deep learning model and a Long Short-Term Memory (LSTM). The LSTM was used to extract sequence information from a sequence of image frames of a specific sign language, while the pretrained deep learning model was used to extract features from single frame images. The dataset used to train the algorithms was gathered in video format from Addis Ababa University. Prior to feeding the obtained dataset to the deep learning models, data preprocessing activities such as cleaning and video to image frame segmentation were conducted. The system was trained, validated, and tested using 80%, 10%, and 10% of the 2475 images created during the preprocessing step. Two pretrained deep learning models, EfficientNetB0 and ResNet50, were used in this investigation, and they attained an accuracy of 72.79%. In terms of precision and f1-score, ResNet50 outperformed EfficientNetB0. For the proposed system, a graphical user interface prototype was created, and the best performing model was chosen and implemented. The proposed system can be utilized as a starting point for other researchers to improve upon, based on the outcomes of the experiment. More high-quality training datasets and high-performance training machines, such as GPU-enabled computers, can be added to the system to improve it.

1. Introduction

Communication is the best tool for interchanging information, ideas, feelings, and opinions to accomplish different tasks and increase the interaction between groups of people. It can be done through verbal (vocal-auditory) and nonverbal (sign-gesture) communication. The verbal communication mechanism is effective to transfer sets of information for intended group, but the nonverbal way is not much effective as the verbal is [1]. Nonverbal way of communication held through gestural movement and visual aid. This type of communication is called sign language [2]. Sign language is one of the natural languages that is most widely used by the hearing and speech impaired (disability) individuals and community. People living with impaired speech and hearing use this gestural communication that is used as a tool for representing their emotions within their groups and common nonimpaired communities [3].

In the world, around 432 million (5.5%) total populations are living with speech and hearing impairment problem [4] and in Ethiopia, a total of 9.9–14.9% of the population are hearing-impaired community [5]. The hearing and speech impairment is either at birth because of prenatal disease or after birth because of illness and accidental noises. Because of their medium of communication, those communities are challenged to get effective

communication when they need service and to work together with unimpaired people. The vast majority of communications technologies are designed to support spoken or verbal and written language (which excludes sign languages), and most hearing people do not know a sign language. As a result, many communication barriers exist for deaf sign language users.

There is a strong need by a deaf community of having a recognizer system for Amharic finger spelling from borrowed alphabets and numeric rather than the all-first Amharic characters. In a previously done research, the authors try to conduct their research in Amharic alphabets from either isolated characters, from only static images, or from character level only. From these gaps, the previous research work lacked applicability of Ethiopian finger spelling in the real world since characters only cannot handle any meaningful format of information. Researchers did not conclude about concepts how borrowed/numeric characters give computer understandable meaning by the use of different recognizer techniques. The aim of this study is, therefore, to investigate and develop a system that recognizes the gesture and sign from video streams/frames of a word level and convert automatically to computer understandable text for effective communication between impaired persons and unimpaired at word level recognition. There is a great contribution about how sign and gestures are identified in addition to the first alphabets and borrowed characters from continuous datasets, applying the suitable techniques for the recognition of EthSL and training a model for real-time detection for Amharic finger spelling for continuous datasets.

Sign language recognition (SLR) is the process of converting signs (gestures) into sequence of texts (alphabet's, words, or sentences) automatically [6]. According to the authors, SLR is the use of machine learning algorithms that can detect the signer (object) gestural position, motion, and hand movements during communication as input and convert these to its equivalent computer understandable format.

Ethiopian sign language (EthSL) recognition is a very recent issue that is developed by different researchers to bridge the gap between those impaired and unimpaired communities. As [7] noted, the recognition system acts as an intermediary between the deaf and normal communities by decreasing barriers of communication. The system they implement is used to translate EthSL based on 34 Amharic alphabets to its equivalent voice format.

Therefore, those communities can communicate easily. As Fantahun and Kumudha [7] mentioned, sign language recognition can be classified into three types. The first one is glove based, in which the recognition system based on sensors at the hand glove of the signer and the sensor speaks about the translation format loudly. The second type of recognition is vision based on the movement of hands and gestures in time series (word and sentence level). This kind of recognition is very challenging to segment each movement into frames and change them to its equivalent text or audio format.

2. Literature Review

2.1. Sign Language. According to [8], sign language is a natural visual-gestural language used as a communication tool for hearing- and speech-impaired communities. It is conveyed through visual and sight facilities, and it has its own grammars and is independent of other verbal and written languages. It has limited vocabularies and makes it not easily communicated and adapted by other common people and very complex as compared to spoken languages.

2.2. Ethiopian Sign Language. EthSL can be used as a sign language communication tool for Ethiopian deaf and hearing-impaired communities. Around 1,000,000 (10%) total populations are hearing- and speech-impaired in the country. Those people living with impairment are very challenged for being educated and communicating with other unimpaired communities, and they do not have sophisticated vocabulary [9]. The sign language that is used by deaf Ethiopian society originated from American Sign Language (ASL) by the influence of Nordic countries (missionaries). Figure 1 presents the Ethiopian finger spelling for EthSL.

The sign language mainly consists of three parts to create a clear understanding about the topic they want to convey. The first one is the use of manual features, which is by the movement of the hand shape and position, the second is the use of nonmanual features, which is by the movement of facials (facial expression), and the third is the use of finger spelling by spelt words or alphabets in their local gesture language [10].

2.3. Ethiopian Finger Spelling. It has another name of Ethiopia manual alphabets, which was developed by Ethiopian association for the deaf (ENAD) in 1971 and latterly accepted by Ethiopian education minster as EthSL. Ethiopian finger spelling was constantly used for representing proper nouns (personal, place, and objects), names with a context/stress based, signs with a country name, technical words, loan signs, and words from a foreign language "since there may not be predefined signs for them" and to represent different terminologies like subjective words [11]. Ethiopian manual alphabets have a total of 34 signs to represent all characters with their correspondence seven movements (derivatives). All the first Ge'ez, characters are represented as a static image without any movement, and the rest (six characters) have their own different position of movement. Those six derivatives are representatives of vowels, whereas the hand configuration is for consonants.

The signer gives a sign by the combination of [12] hand shape, hand orientation, and hand movements.

2.3.1. Hand Shape. The proper and exact image for the intended message can be described based on the right-hand shape of the signer. Each sign for characters (alphabets),

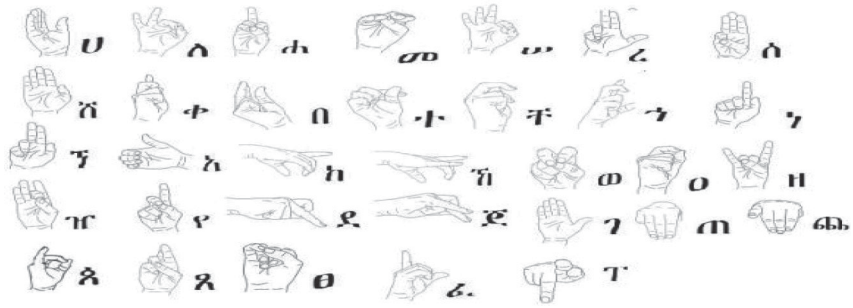

Figure 1: Ethiopian finger spelling for EthSL [9].

numbers, words, and sentences has a different hand shape representation as shown in Figure 2.

2.3.2. Hand Orientation. It refers to the palm position in which the signer uses to represent something. Different orientations also represent different meanings; if the orientation changes it, the meaning also changes. It can include the palm facing in, out, horizontal, left, up, or down. The same characters of preprocessed images with various hand orientations are shown in Figure 3.

2.3.3. Hand Location. The location of the hand can move to different directions to represent messages. Mostly, they used locations like around the forehead, chest, left to right, right to left, and some sinusoidal movements of the hand. As shown in Figure 4, there are many more images for different hand locations for representing derived letters.

2.3.4. Hand Movement. Some signs have static images, and some have signs with motion. Knowing the proper movement of the hand, palm, and shoulder is very important for getting what we want to convey. Table 1 defines a variety of trajectories of Amharic letter "" to give another derived letter of the family name.

2.3.5. Nonmanual Features. In addition to the hand position, movement, and location, the use of facial expression is included to clear the feelings and degree behind the word. This is like expressing the degree of happiness surprise and sadness.

2.4. Deep Learning. Machine learning (ML) is used for complex and new tasks that are very tedious for manual work by generating algorithms based on rules or relationship between data sets and recognizing patterns in the given data. Deep learning is a special kind of ML algorithm, which works based on artificial neural networks (ANN) by taking inputs from artificial neurons and adjusting them based on the pattern they designed. The existence of deep learning is due to its data dependency rather than other ML algorithms since deep learning works with huge amount of data for perfect prediction, but other ML algorithms perform well with small amount of data. Deep learning is more special for high problem-solving approach and its high execution time [13].

2.5. Supervised Learning and Unsupervised Learning. Supervised learning is a category of machine learning that is used for the classification, categorization, and recognition of objects through the use of labeled dataset to train an algorithm. This learning category is mainly used for the classification and any of prediction tasks by a machine learning [14]. According to IBM [14], supervised learning can be used for solving problems like classification and regression. The classification operation is done through the use of algorithms, whereas unsupervised classification is most important in the process of expressing and clustering for the classification based on their similarities of data rather than prediction of classification. It involves the separation and expression of data into its groups based on their similar characteristics [15]. Here, for unsupervised learning operation, the algorithms are used for finding hidden structures and intrinsic pattern in the use of unlabeled data inserted. Some of the algorithms used for unsupervised learning are K-means clustering, anomaly detection, K-nearest neighbors, and association.

3. Related Works

Rasha and Muntadher [16] attempted to develop a real-time SLR system for ASL by including all ASL finger spelling. The recognizer system was built based on the use of CNN and very deep neural networks (VGG-NET) for the multiclass classification of 28 classes with the total of 61, 614 row input images. The recording and capturing of inputs were dependent on the real coloring images. As a result, an accuracy of 99.67% for training, 99% for testing, and 100% validation rate was obtained, and the use of VGG-NET gave them a better improvement in their system performance than previous efforts.

Sugandhi and Sanmeet [17] conducted an experiment to help hearing-impaired persons to effectively communicate with other normal Indian community. The recognition system was based on the Indian sign language of numbers (1–9), which are collected in the form of videos by the resolution of 64×64 pixels. The researcher used CNN algorithm with two hidden layer and used a dimensionality reduction of max pooling to reduce overfitting and training

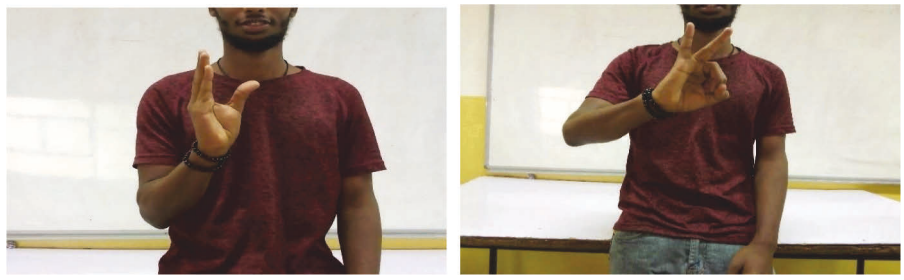

Figure 2: Sign representation of two different characters (ሀ and ለ, respectively).

Figure 3: Different hand orientation for the Amharic character.

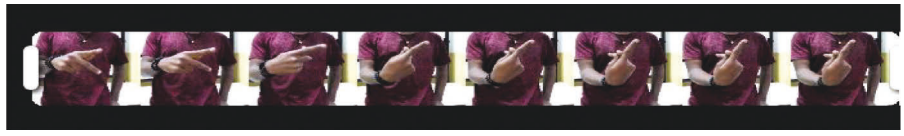

Figure 4: Hand locations for representing a letter derived from.

Table 1: Different trajectories of the letter "" with all types of motion [9].

Order	2nd	3rd	4th	5th	6th	7th
Form						
Trajectory						
Direction	Left	Right	Down	Nearly circle	Down oscillatory	Rotation

time. The data set was trained with Adam optimizer at 10 epochs with the total accuracy rate of 99.56%.

Legesse [18] attempted to develop Ethiopian finger spelling classification towards automating Ethiopian sign language. The study was limited to finger spelling with static signs while capturing the input image. As the authors discussed, this research has contributions to teleconferencing, videoconferencing, and increasing the quality of life for the community. The overall accuracy was 88.08% for using NN-principal component analysis (PCA) feature, 96.22% NN-harr like feature, and 51.44% by template matching.

Nigus [9] attempted to develop Amharic sign language recognition based on Amharic alphabet signs from the Amharic first characters. The significance for the author's contribution was to address special needs for the disabled community specially for hearing-impaired ones. This research mainly focused on reducing the communication gap between impaired and unimpaired communities, to support students to learn Amharic alphabet signs quickly and to help another researcher with the corpus when it is collected during conducting his research. The methodology of algorithm used by the researcher was support vector machine (SVM) and neural network (NN) by giving 90% of the collected image for training model and 10% of test model. The overall accuracy of the recognizer system was 57.82% of the NN and 74.06% by SVM.

Isayas and Hussien [19] aimed to develop Amharic sign language recognition model for Amharic characters using deep learning approach. The author aimed at recognizing Amharic selected alphabets; those are the first of the tenth. The methodology of the research was the use of CNN and faster R-CNN model for training and testing of the system.

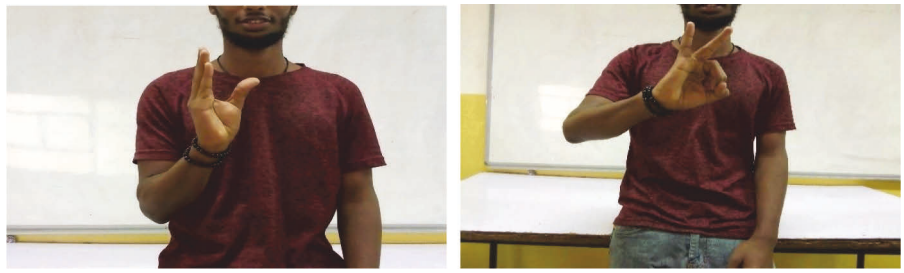

Figure 5: Signs for different Amharic character.

Table 2: Amharic words to represent the number of classes for the model construction.

Classes	Data		
	Training	Testing	Validation
/hana	84	14	14
/hawassa	120	20	20
/lema	96	16	16
/car	144	24	24
/honey	96	16	16
/sehar	138	23	23
/sara	108	18	18
/kebena	138	23	23
/kana	108	18	18
/augest	138	23	23
/jimma	138	23	23
/x-mass	120	20	20
/gari	120	20	20
/shoes	138	23	23
/horse	144	24	24
Total	1865	305	305

Overall, the total system accuracy was 98.25% for single shot detector (SSD) and 96% for faster R-CNN. For future research, the authors outline that, to use dynamic/different background while capturing input video, we take into consideration using all Amharic characters to fully recognize a system to develop for word/sentence level and to develop a system for real-time recognition.

4. Methodology

Construction of Amharic sign language for Ethiopian finger spelling is done based on the input dataset of both images and videos. Those datasets in the form of videos are changed into sequence of frames to be ready for the feature extraction tasks. The developed recognizer model was followed by an experimental research design. The design was held based on utilizing of approaches necessary for a computer vision tasks. The study used image processing and supervised learning algorithms for the sign language recognition task.

4.1. Data Collection. Data for the research was collected from the signer based on primary sources in Addis Ababa University, 6-kilo campus, and from Ethiopian sign language and deaf culture studies, the department of linguistics and from Ethiopian center for disability and development (ECDD). The sampling technique used by a researcher for a data collection was nonprobabilistic convenience sampling, which works with only with the voluntarily participants of the deaf person. The total amount for recorded data was including all Amharic alphabets from the first to the seventh numeric characters and two borrowed characters. Different signs for different Amharic characters are presented in Figure 5.

The total amount of classes for images and videos was 240 and total 26,424 extracted frames. From the collected number of Amharic characters, fifteen different words were randomly selected for the research demonstration with 2475 images of dataset. From 2475 images, 80% was given to the training, 10% for validation, and the remaining 10% for the testing sets. Table 2 shows randomly selected Amharic words to be recognized by a classifier model as a final work.

4.2. Dataset Preparation. After changing collected video into a sequence of framed images (video processing), there is an image preprocessing task (cropping, resizing, and normalization) to eliminate noises during the data capturing. During the capturing of datasets, the image and video were in huge size and contained unnecessary objects from the background, and they still need cropping. Frames with huge size need a greater computational space; as a result, the height and width of the frame need cropping. Frames from the original size of 3264×2448 were resized to 2000×1200, and frames from 1280×720 resized to 500×400 by using if condition in python programming.

RGB colored images were also changed to BGR, by the use of CvtColor conversion in python method, and we got a great effect on the images. CvtColor method is used to change the color of an image from one color space to another. The reason of changing RGB to BGR is an OpenCV function for reading of an image file that needs a color space arrangement of BGR. The images presented in Figure 6 are sample examples for the color conversion from RGB to BGR.

4.3. Feature Extraction. A CNN is a type of NN designed to extract features from data to classify them into their representatives and reduced vector form [20]. Transforming the input data into the set of features is called feature extraction. Features in computing refer to the fact that the variables and attributes represent a quantifiable property of an object in object detection, classification, and recognition. The hand shape, motion, and color can be some of the basic features extracted from the image frame. For the

Figure 6: A cropped image of RGB color VS BGR color.

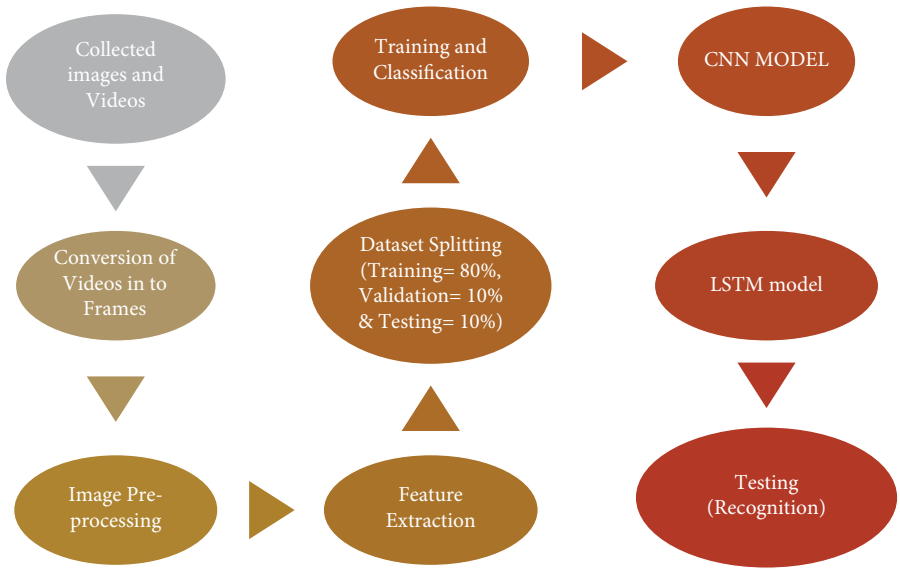

Figure 7: The proposed Ethiopian finger spelling recognizer model's system architecture [9].

representation of shapes, contour-based and region-based shapes are used. Features are extracted from the contour only in the case of contour-based classification and extracted from the whole shape region for the region-based classification [11].

The possible features extracted for the model construction are done through the use of hybrid CNN-LSTM and a pretrained efficientNetB0 and Resent50 as a feature extractor network architecture. Both algorithms are of a CNN architecture and allow an image to forward-propagate to the final max-pool layer. The output of the max-pooling layer for efficientnetB0 has a shape volume of $7 \times 7 \times 1280$. The next step was the process of incremental learning, which plays a great role in deep learning to extract features on relatively large datasets. This is used to keep the model from taking all the entire dataset at once for a training by fitting to the available RAM. This is not essentially tolerable, because the datasets may have and take huge amount of memory. So, what we did is an incremental learning of selecting a batch of data to learn at once during the training phase. We train the model with $224 \times 224 \times 3$ and $150 \times 150 \times 3$ input shapes by efficient netB0 and resnet50, respectively, and a batch size of 50. Figure 7 presents the architecture of a proposed Ethiopian finger spelling recognizer model.

5. Results and Discussion

The model constructed is now ready to detect Amharic selected words that were constructed from the combination of different Amharic Alphabets. The trained (efficientNetB0 and ResNet50) classifiers are fitting with a data of an image with the portioning of training and testing sets. The testing and training work on validation and training of data sets with batch size 50 from the total 2475 images and 10 number of epochs for the entire training phase. The need for the batch size was because of the huge number of datasets. The lower number of batch size is needed to increase steps per epoch for iteration and bring a good accuracy result. During the training of a model, the accuracy rate is telling how well the model performs, and the loss is the result for bad predictions.

5.1. Test Results of EFFICIENTNETB0. From the 1865 training datasets, the model fits with a batch size of 50 and epochs of 10. The model brings overall accuracy of 72.79%.

```
Epoch 1/10
38/38 [==============================] - 130s 3s/step - loss: 0.6289 - accuracy: 0.6858 - val_loss: 0.6080 - val_accuracy: 0.7115
Epoch 2/10
38/38 [==============================] - 124s 3s/step - loss: 0.6456 - accuracy: 0.6740 - val_loss: 0.5869 - val_accuracy: 0.7279
Epoch 3/10
38/38 [==============================] - 124s 3s/step - loss: 0.6723 - accuracy: 0.6579 - val_loss: 0.5963 - val_accuracy: 0.7279
Epoch 4/10
38/38 [==============================] - 124s 3s/step - loss: 0.6242 - accuracy: 0.6777 - val_loss: 0.5806 - val_accuracy: 0.7279
Epoch 5/10
38/38 [==============================] - 124s 3s/step - loss: 0.6004 - accuracy: 0.6960 - val_loss: 0.6002 - val_accuracy: 0.7279
Epoch 6/10
38/38 [==============================] - 123s 3s/step - loss: 0.6177 - accuracy: 0.6810 - val_loss: 0.5766 - val_accuracy: 0.7213
Epoch 7/10
38/38 [==============================] - 123s 3s/step - loss: 0.6183 - accuracy: 0.6772 - val_loss: 0.5620 - val_accuracy: 0.7279
Epoch 8/10
38/38 [==============================] - 123s 3s/step - loss: 0.6382 - accuracy: 0.6783 - val_loss: 0.5768 - val_accuracy: 0.7279
Epoch 9/10
38/38 [==============================] - 122s 3s/step - loss: 0.6213 - accuracy: 0.6761 - val_loss: 0.5621 - val_accuracy: 0.7016
Epoch 10/10
38/38 [==============================] - 120s 3s/step - loss: 0.6059 - accuracy: 0.6810 - val_loss: 0.5861 - val_accuracy: 0.7279
```

Figure 8: Overall accuracy of EfficientNetB0 model.

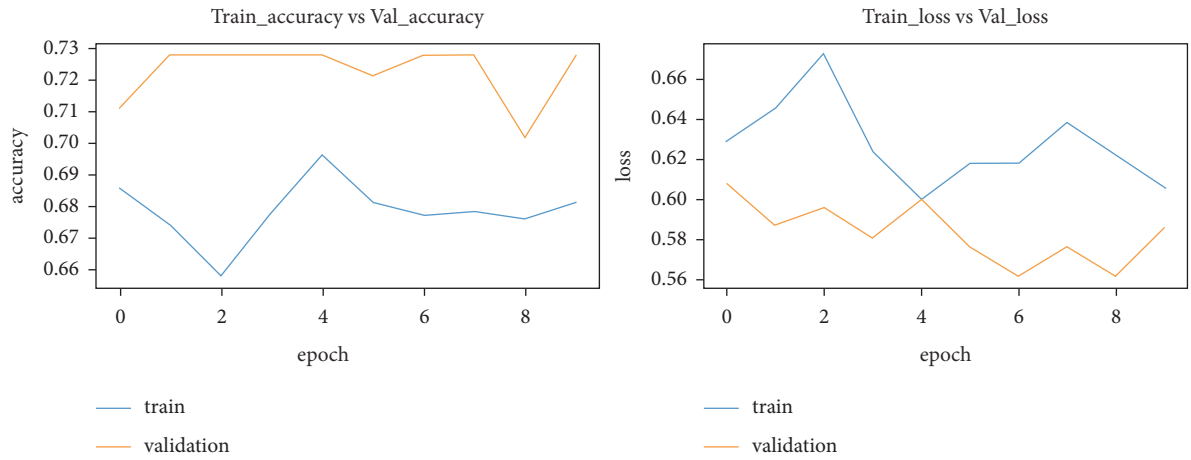

Figure 9: Training accuracy and val_acc Vs training loss and val_loss of EfficientNetB0.

```
Epoch 1/10
38/38 [==============================] - 124s 3s/step - loss: 0.7080 - accuracy: 0.6735 - val_loss: 0.6159 - val_accuracy: 0.7279
Epoch 2/10
38/38 [==============================] - 121s 3s/step - loss: 0.6841 - accuracy: 0.6676 - val_loss: 0.6052 - val_accuracy: 0.7279
Epoch 3/10
38/38 [==============================] - 121s 3s/step - loss: 0.6875 - accuracy: 0.6767 - val_loss: 0.6038 - val_accuracy: 0.7213
Epoch 4/10
38/38 [==============================] - 124s 3s/step - loss: 0.6577 - accuracy: 0.6735 - val_loss: 0.6435 - val_accuracy: 0.7213
Epoch 5/10
38/38 [==============================] - 122s 3s/step - loss: 0.6586 - accuracy: 0.6767 - val_loss: 0.7537 - val_accuracy: 0.6623
Epoch 6/10
38/38 [==============================] - 121s 3s/step - loss: 0.7244 - accuracy: 0.6665 - val_loss: 0.6462 - val_accuracy: 0.7148
Epoch 7/10
38/38 [==============================] - 121s 3s/step - loss: 0.6684 - accuracy: 0.6745 - val_loss: 0.6859 - val_accuracy: 0.6951
Epoch 8/10
38/38 [==============================] - 121s 3s/step - loss: 0.6449 - accuracy: 0.6847 - val_loss: 0.5841 - val_accuracy: 0.7246
Epoch 9/10
38/38 [==============================] - 123s 3s/step - loss: 0.6425 - accuracy: 0.6815 - val_loss: 0.5980 - val_accuracy: 0.7279
Epoch 10/10
38/38 [==============================] - 124s 3s/step - loss: 0.6324 - accuracy: 0.6799 - val_loss: 0.5761 - val_accuracy: 0.7279
```

Figure 10: Overall accuracy of ResNet50 model.

Each step takes a minimum of three seconds and an average of 124 microseconds for each step per epoch as shown in Figure 8.

The developed model starts increasing the validation accuracy while decreasing the validation loss at the beginning of training time in a model that is built with better learning, and it works fine, but at the end of the training, validation accuracy increases with an increasing rate of validation loss. This tells us that the model is in case of overfitting. The validation accuracy of the model decreases when the epochs increased at the end of the epochs, and this represents that the model is fitting with the training set better but losing the ability to predict new data and starting overfitting as shown in Figure 9.

5.2. Test Results of RESNET50. With the same training dataset, the number of batch sizes and the number of epochs with EfficienNetB0 and ResNet50 models bring the same overall accuracy of 72.79%. It has an average of 122 microseconds for each step per epoch. The time taken is relatively small as compared to EfficienNetB0 for finishing the training task and having higher learning rate. Figure 10 shows the developed ResNet50 model performance.

Figure 11 illustrates that throughout the initial stages of the model training, validation accuracy rises as training accuracy falls, and validation accuracy remains constant while training accuracy falls. As a result, the model has likely stopped learning and is overfitting. In summary, when validation accuracy rises, the model validation loss begins to rise. This is as a result of the developed model being completely overfitted.

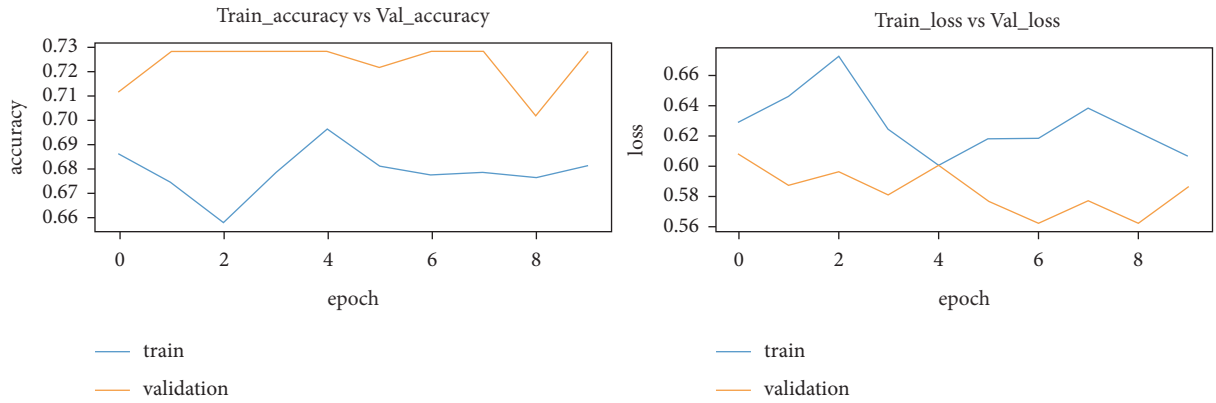

Figure 11: Train_acc and Val_acc Vs Train_loss and Val_loss for ResNet50.

Table 3: Results on the number of recognized images using EfficientNetB0 and ResNet50 models.

No_of test signs	Recognition in number	Recognition percentage (%)
Correctly recognized signs	222	72.79
Incorrectly recognized signs	83	27
Total	305	

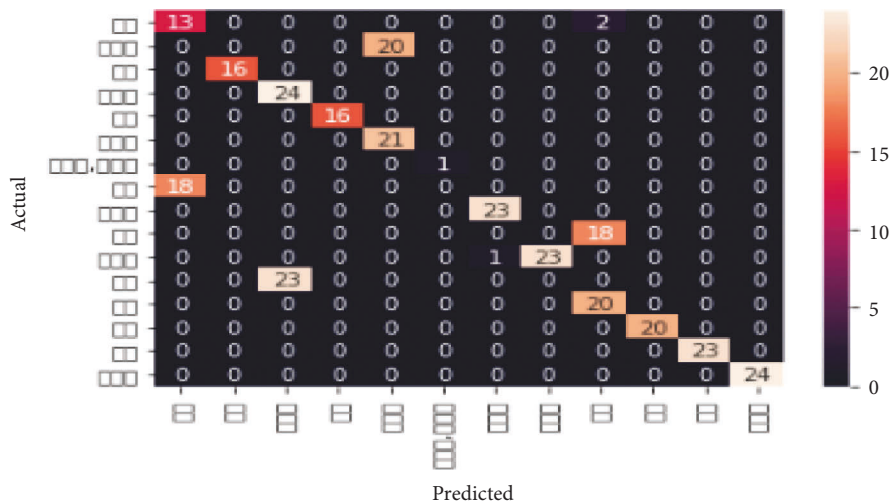

Figure 12: Confusion matrix for developed recognition model.

Although the model virtually accurately categorizes more than half of the testing dataset even when it is overfit, the training data still must be balanced by some dataset organization. Table 3 shows the total number of images being recognized correctly and incorrectly from the total images of 305.

5.3. Comparison of the Classifier Performance. The accuracy of the model was obtained by the ration of adding all correctly classified true positive values in the diagonal line and all the values above and below the diagonal values that represent the falsely classified positive and negative values. Figure 12 represents the total model performance by a confusion matrix.

$$\text{Accuracy} = \frac{13 + 16 + 24 + 16 + 21 + 1 + 23 + 23 + 20 + 20 + 23 + 24}{224 + 2 + 20 + 18 + 18 + 23} = 0.73. \quad (1)$$

From Figure 12, an approximate 73% of accuracy was obtained. The recall for the model is higher than that of precision and f1-score, which tells us that the model created correctly classified the predicted values and the more true positive values in the model. The precision of the model is relatively low as compared to recall and f1-score, and the

	precision	recall	f1-score	support
ሀኘ	0.42	0.87	0.57	15
ሀሞሳ	0.00	0.00	0.00	20
ሰማ	1.00	1.00	1.00	16
መኪና	0.51	1.00	0.68	24
ማር	1.00	1.00	1.00	16
ሰሃር	0.51	1.00	0.68	21
ሳራ	0.00	0.00	0.00	18
ቀበና	0.96	1.00	0.98	23
ቃና	0.00	0.00	0.00	18
ነሃሴ	1.00	0.96	0.98	24
ጅማ	0.00	0.00	0.00	23
ገና	0.50	1.00	0.67	20
ጋሪ	1.00	1.00	1.00	20
ጫማ	1.00	1.00	1.00	23
ፈረስ	1.00	1.00	1.00	24
micro avg	0.73	0.73	0.73	305
macro avg	0.59	0.72	0.64	305
weighted avg	0.60	0.73	0.65	305

Figure 13: Performance metrics for the developed model.

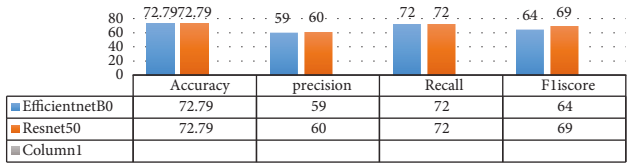

Figure 14: Model comparison by four different performance metrics.

Figure 15: An image for a real-time detection feeding.

Figure 16: Image of real-time detection for Amharic word "መኪና".

model has higher false positive values. The harmonic mean of a model (f1-score) is about the model that has the lower number of false rate values (FP and FN). Figure 13 shows the performance metrics for the developed model.

The test results for different models with the same dataset are shown, as well as the number of batch sizes and epochs that unfortunately scored 72.79%, but the use of ResNet50 has fast learning rate, higher precision, and higher f1-score. The higher the precision of the ResNet50 model means, the lower the rate of false positive values in the classifier model. A high F1-score value was scored in ResNet50 as well, which has a relatively low number of false negative and false positive values than those of EfficientNetB0. We are capable of correctly classifying real threats through the use of ResNet50 and make it perform better than EfficientNetB0. Figure 14 shows the final model performance comparison of constructed model.

Signs that are not correctly classified for the two models are four out of fifteen number of words. The false classified words are , , , and . The incorrect classification of the words is because the alphabets "" with "," "" with "," "" with "" and "" with "" have little similar arrangements. This makes Amharic alphabet recognition system performance lower than that of other language alphabets.

5.4. Real Time Detection. The final stage is to create a real-time detection model, as shown in Figures 15 and 16. The real-time detection is then performed using the previously trained and saved. h5 model. The ability to recognize objects in real time is utilized to include and make information available to a community group. Because most live communications are based on textual languages, it will also be utilized to facilitate live communications such as video conferencing.

An OpenCV detects object features (shapes, edges, and positions of a hand movement) as an input for detection and recognition in real time, as shown in Figure 15.

In the image, a recognizer model tries in real time to recognize an Amharic word "" presented in green text format and begins to collect all of the object's necessary movements. The built model is run and provides real-time data on what is happening in the environment. The developed real-time model is used to process streaming data and has the capacity to sense, evaluate, and act on streaming data as it comes in without having to ingest and store it in a back-end database. To process streaming data asynchronously, real-time applications frequently use event-driven architecture. It also enables real-time programs to respond to user needs faster and more efficiently than the manual sign language translation by experts. Figure 16 shows a real-time recognition for selected

Amharic words from the screenshot of a model in live detection.

6. Conclusion

There are millions of hearing-impaired people in Ethiopia who are unable to communicate due to communication channel failures. Hearing people are not experts at deciphering sign language, and deaf people are not experts at deciphering written and spoken languages. The established recognizer model will have a significant impact on bridging the gap in their communication. The proposed model takes a sign language from a video file and converts it into numerous sequential frames in order to translate it into Amharic text. Image preprocessing, feature extraction, and classification are the three basic components of the model. The former employs a variety of image processing approaches, while the second and third employ CNN and LSTM classifiers to complete the specified task using EfficientNetB0 and ResNet50 classifiers, respectively. From the EfficientNetB0 and ResNet50 models, the model achieves an overall accuracy of 72.79 percent.

7. Recommendations

This research has made a significant contribution to the Ethiopian sign language by bridging a significant communication gap between the deaf and hearing communities. Taking into account the following notions, we offer several feature works for coming researchers:

(i) Long words and phrases should be added to the sign language recognition model, and data from various users should be collected to better comprehend signers' hand arrangement.

(ii) We virtually always employed the same background during data collecting, and future studies should take into account varying backgrounds from the real world.

(iii) Because this study only focuses on one-way communication (sign language to text), future studies should think about recognizing sign language to text in a harmonic method of sounds to help blind and deaf people.

(iv) Finally, the researchers should focus on establishing an Android application system that can recognize and translate Amharic sign languages into a computer-readable format.

Data Availability

The datasets (images and videos) used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

During the conduction of this thesis, the following are the contributions of the author. (i) Searching out the communication gap of a deaf community. (ii) Collecting and preparing well organized corpus for the model construction. (iii) The application of deep learning concepts which is state-of-the-art to solve local problems. (iv) Representation and recognition of Amharic sign language into a text for bridging the gap of communication between deaf and hearing people.

Acknowledgments

Thanks and appreciation are due to Dr. Andargachew and Mr. Kaleab for their unlimited support during the data collection. They also thank Jimma University for the funding (25,000 ETB) for conducting this thesis and the Faculty of Computing and Informatics for all the opportunities to stay there.

References

[1] B.-N. Liubov, "Verbal communication skills," 2015, https://www.researchgate.net/publication/281784451.

[2] F. Lunenberg, "Communication: the process, barriers and improving effectiveness," *American Journal of Industrial and Business Management*, vol. 6, 2010.

[3] R. Sabeenian, B. Sai, and A. Mohamed, "Sign language recognition using deep learning and computer vision," *Journal of Advanced Research in Dynamical and Control systems*, vol. 12, no. 5, pp. 1–9, 2020.

[4] DTE Staff, "30 years down to earth," 2021, https://www.downtoearth.org.in/news/health/every-4th-person-to-suffer-hearing-loss-by-2050-who-75718.

[5] E. Yigremachew and W. Endashaw, "A real-time Ethiopian sign language to audio converter," *International Journal of Engineering Research in Africa*, vol. 8, no. 8, pp. 1–6, 2019.

[6] D. E. Philippe, "Spoken language processing techniques for sign language recognition and translation," 2008, https://www-i6.informatik.rwth-aachen.de/.

[7] A. Fantahun and R. Kumudha, "Ethiopian sign language recognition using artificial neural network," in *Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications*, Cairo, Egypt, November 2010.

[8] K. G. Ashok and K. R. Kiran, "Sign language recognition: state of the art," *Asian Research Publishing Network Journal of Engineering and Applied Science*, vol. 9, no. 2, pp. 1–8, 2014.

[9] K. Nigus, "Amharic sign language recognition based on Amharic alphabet signs," M.Sc Thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2018.

[10] C. Helen, H. Brian, and B. Richard, "Sign language recognition," in *Visual Analysis of Humans: Looking at People*, M. Thomas and H. Adrian, Eds., Springer, Berlin, Germany, pp. 539–562, 2011.

[11] G. Tefera, "Recognition of isolated signs in Ethiopian sign language," M.Sc Thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2014.

[12] K. Šarac and A. Ninoslava, "Phonological parameters in Croatian sign language," *Sign Language and Linguistics*, vol. 9, no. 1, pp. 33–70, 2006.

[13] C. E. Raffaele, "Artificial intelligence and machine learning applications in smart production: progress, trends and directions," *Sustainablity*, vol. 12, 2020.

[14] IBM Cloud Education, "IBM cloud learn hug," 2020, https://www.ibm.com/cloud/learn/unsupervised-learning.

[15] D. EnNovemberderle and R. Weih, "Integrating supervised and unsupervised classification methods to develop a more lop a more accure accurate land co ate land cover classification er classification," *Journal of the Arkansas Academy of Science*, vol. 1–9, 2005.

[16] A. Rasha and K. Muntadher, "A real time American sign language recognition system using convolutional neural network for real dataset," *TEM Journal*, vol. 9, no. 3, pp. 937–943, 2020.

[17] P. Sugandhi and K. Sanmeet, "Sign language generation system based on Indian sign language grammar," *ACM Journals, ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 19, no. 4, pp. 1–26, 2020.

[18] Z. Legesse, "Ethiopian finger spelling classification: a study to automate Ethiopian sign language," Addis Ababa University, Addis Ababa, Ethiopia, Master of Science, 2008.

[19] F. Isayas and S. Hussien, "Developing amharic sign language recognition model for amharic characters using deep learning approach," Rresearch Square, vol. 1, pp. 1–5, 2021.

[20] L. Matilda, "Feature extraction for image selection using machine learning," Linköping University, Linköping, Sweden, Masters of Science, 2017.