
Ayodele Olabanji, Akinlolu Adediran Ponnle

Abstract — Sign language is the primary method of communication adopted by deaf and hearing-impaired individuals. The indigenous sign language in Nigeria is one area receiving growing interest, with the major challenge faced is communication between signers and non-signers. Recent advancements in computer vision and deep learning neural networks (DLNN) have led to the exploration of necessary technological concepts towards tackling existing challenges. One area with extensive impact from the use of DLNN is the interpretation of hand signs. This study presents an interpretation system for the indigenous sign language in Nigeria. The methodology comprises three key phases: dataset creation, computer vision techniques, and deep learning model development. A multi-class Convolutional Neural Network (CNN) is designed to train and interpret the indigenous signs in Nigeria. The model is evaluated using a custom-built dataset of some selected indigenous words comprising of 15000 image samples. The experimental outcome shows excellent performance from the interpretation system, with accuracy attaining 95.67%.

Index Terms — Sign Language, Convolutional Neural Network, Interpretation System, Indigenous Sign Language.

I. INTRODUCTION

Sign language is the most natural and raw system of languages, which predates the early advent of human evolution. The evolution of sign languages follows the early theories in history, with appearance starting even before the emergence of spoken languages. Over time, sign language has been adopted and evolved into an integral part of the daily communication process. Currently, sign languages are extensively applied in international sign use for religious practices, in the sports world, and in workplaces [1].

Sign Language remains an ongoing research study since the arrival of computer vision techniques and has recorded notable progress in the use of several advanced techniques. This progress has seen sign language's identification move from being a not possible research area to one with a growing level of interest. The goal of the sign language interpretation system is to deliver a suitable platform that translates known sign into text and provide support to conversations between signers and non-signers. Besides, the interpretation system plays a critical role in human-computer interaction and graphical applications [2].

In recognition of sign language, the process of determining precise figures and shapes still faces some degree of uncertainty. This uncertainty can be attributed to several factors, including high clutter, image scale, occlusion, low lighting, and more. These factors hinder the possibility of attaining an accuracy level of 100%.

Indigenous sign language in Nigeria is the most popular communication approach for individuals with difficulties in speaking. This approach is used to express emotions and communicate efficiently. However, only a few non signers use or understand the indigenous sign language used in Nigeria. The alternative usage of written communication is considered inconvenient, as the deaf community is generally less skilled in writing the vocal language [3]. Hence, there is a need for the integration of an interpretation system into the communication methods.

In this paper, we present the development of a computer-aided interpretation system for indigenous sign language in Nigeria. The system created helps signers interpret indigenous signs of indigenous words towards communicating with non-signers. A new image dataset for some indigenous sign language for some indigenous words in Nigeria was created. The approach used takes in video sequences and extract multiple frames. The final dataset comprises image samples further processed using image processing algorithms comprising the Hue, Saturation and Value (HSV) thresholding; Histogram Back-projection; Otsu; Gaussian filter; and Median Filter. The dataset is used to train the Convolutional Neural Network (CNN) for feature extraction and classification purposes. The interpretation system was deployed using an interactive graphical user interface to engage users.

II. RELATED WORKS

Multiple research endeavors to create an optimal algorithm have led to the creation of a number of techniques. For example, some techniques developed for the interpretation of various known sign languages include Albanian sign language [4]. American sign language [5], Arabic Sign Language [6], [7]; Chinese sign language [8], Korean sign language [9], Mexican sign language [10], and Tamil sign language [11].

Submitted on June 03, 2021.
Published on June 28, 2021.
A. O. Olabanji, Department of Electrical and Electronics Engineering, Federal University of Technology, Akure, Nigeria.
(corresponding e-mail: olabanjiayodele5@gmail.com)

DOI: http://dx.doi.org/10.24018/ejce.2021.5.3.332
A new technique created by [12] entails YCbCr (Luminance; Chroma: Blue; Chroma: Red) conversion, edge detection, shape enhancement, and CbCr (Chroma: Blue; Chroma: Red) mapping. [13] identified a new technique that combines the TSL (Tint, Saturation, Luminance) skin color model and the improved Kalman filter. The technique applied limits any extra skin color objects through the improved Kalman filter model, which evaluates the motion and its midpoint location. The use of gloves by [14] employs random distributions of 10 full saturated colors in 20 patches. Nevertheless, the need for regular tuning would robustly enhance its operation [15], [16].

Fuzzy C-Means clustering (FCM) and thresholding technique was employed by [17] for segmenting the face and hand. The FCM uses a clustering method that applies an iterative approach to engage fuzzy partitioning. The Fuzzy C-Means algorithm first clusters the image using its color data. The mean value is finally selected by thresholding that uses the possible values available to the skin tone. [18] applied a method involving two-step hand segmentation. The method applies canny edge detection on the images first to identify the edges. The second phase involves applying the seeded region growing technique to segment the hand from its background.

[19] use morphological operations and HSV (Hue, Saturation, Value) thresholding to segment the hand from its background. Determination of the HSV threshold values is done based on the skin color of the hand before applying it to the images. Morphological operations, including dilation and erosion, are further applied to eliminate noise. [20] used three-hand segmentation algorithms in the development of a recognition system. The three algorithms include the Otsu algorithm, thresholding in YCbCr color space, and the Gaussian Mixture Model (GMM).

[21] employed a heuristic optimization algorithm for the hypermeter optimization process. The hyperparameters for the CNN structure used are based on the AlexNet model, alongside the outcome from the optimization algorithms. A 98.09% and 98.40% recognition accuracy were achieved on two public datasets used for evaluation. [22] used connected component analysis algorithm on the hand to segment the tip of the fingers as its Region of Interest (RoI). Gesture detection and recognition model using the segmented finger was developed via CNN, which achieved a 96.2% recognition accuracy.

Over the past years, several techniques have been developed for different sign languages used in different countries. For the indigenous sign language of indigenous words in Nigeria, which is possibly adopted and adapted from the ASL, the interpretation system presented is designed to solve the peculiarities concerning indigenous words in Nigeria.

III. BACKGROUND STUDIES

Most recognition systems focus on three major techniques for hand sign analysis: appearance-based, depth-based, and sensor-based systems. The appearance-based analysis of hand signs combines two key stages:

i. Detection/Tracking;
ii. Recognition.

A. Detection/Tracking

Detection plays a key part in the interpretation system. Selective preprocessing and segmentation are required mainly on the raw input samples to aid feature extraction from random/uniform backgrounds. Tracking techniques involve reading the hand movement and its position [23]. Notable steps to the detection of the hands are:

i. Color Space Thresholding;
ii. Binarization;
iii. Background Removal.

B. Recognition

The primary task for an interpretation system is to identify hands-signs correctly for specific positions. Dynamic and static hand signs are two recognition techniques in the research of hand sign movement.

IV. DATA PROCESSING

Fig. 1 shows the framework for the proposed interpretation system. The interpretation systems for sign language combine three key stages: Segmentation, Feature Extraction, and Classification. The segmentation process is characterized based on the capture system employed and focuses on obtaining the hand shape from input streams. The hand tracking and detection process uses the appearance-based method.
A. Data Collection

To create the database for the interpretation system, the selected indigenous signs and their connotations were obtained, with each containing several samples. The study employed the use of questionnaire, physical interaction, and indigenous sign illustrations [24]. The appearance-based technique was employed to acquire the needed data. This method was chosen because it is user-friendly and a popular option in sign recognition systems.

RGB cameras were used to capture multiple image frames of the hand signs needed to create the dataset. The use of multiple samples per sign collected helps increase the accuracy of the system. The collected signs were adequate to create both the training, validation, and test dataset for the system. The image data and labels are saved into an SQLite database. The collected data served as direct inputs for the machine learning algorithms. The word dataset contains 15 words from the indigenous sign language in Nigeria. Fig. 2 shows some samples from the created dataset. The word dataset has the following features:

i. Each class has 1000 image samples from several signers and is saved in the .JPG format.
ii. The total image samples contained in the dataset are 15000 (1000 * 15) images.

B. Preprocessing

From the raw hand-sign samples with normal backgrounds, the hand area is cropped out from the images. The approach applied focuses on the hand, the key element needed to help the interpretation system. The system converts the image samples in RGB (Red, Green, Blue) to HSV color space. The images are converted into binarized colour images using the histogram back-projection and OTSU approach [25] as the automatic thresholding technique. All images obtained are further re-sized by 50×50-pixel size.

C. Augmentation

The use of real-world data usually features the tendency to be random and offers several distortion types, including rotation, shifting, and more. Data augmentation remains a popular strategy [26] used to enhance the performance (with regards to accuracy) of a particular dataset. To enhance the dataset with such situations, image augmentation techniques are applied. The image samples are rotated randomly using 0 and 10 degrees, both in clockwise and anticlockwise directions. The augmentation applied to the system data shows promising results. Fig. 3 shows samples of the augmented images.

D. Splitting Dataset

The word dataset contains 1000 exemplars for each of its 15 signs. The total number of images in the dataset is 15000. The word dataset is further split into training, validation, and test set. The three sets are split using the ratio of 80:10:10, respectively.

V. DEEP LEARNING MODEL

A deep neural network is engaged in this system to interpret indigenous hand signs. The neural network used is Convolutional Neural Network. The CNN model used, model training, evaluation, and other methodologies are discussed in this section.

A. Model Design

The system uses the created dataset as input to the CNN to classify the indigenous sign language. The CNN is able to perform feature extraction on the dataset. The CNN implementation and model training process were achieved using the Keras library. The CNN architecture comprises of the following layers: 2D Convolutional Layer, Max-pooling Layer, Dense Layer, and Output Layer. A small Region of Interest (RoI) was employed for the image samples to avoid disingenuous performances and overfitting in the learning process.

The CNN model comprises three convolutional layers, two pooling layers, two fully connected layers, combining the ReLU (Rectified Linear Unit) and SoftMax. In addition, the batch-normalization [27] learning approach is employed to avert overfitting. The first convolutional layer takes in the input shape using 8 kernels of size $1 \times 1$, followed by a batch normalization layer. Convolution is centered around merging two sets of information, and in the case of CNN, they are the Kernel function and input data. Considering a kernel $K[i,j]$ with dimensions $[2I + 1, 2J + 1]$ where $(i,j) = (0,0)$ is the center of the kernel, the output $g[m,n]$ applying the kernel function on the input $a[m,n]$ is given in (1) and (2).

![Fig. 2. Samples from the created indigenous sign language word in Nigeria.](image1)

![Fig. 3. Augmented samples from the word dataset.](image2)
\[ g[m, n] = K[i, j] \ast a[m, n] \]  
\[ g[m, n] = \sum_{i=-1}^{1} \sum_{j=-1}^{1} K[i, j] \cdot a[m - i, n - j] \]

The second convolutional layers follow with 16 kernels of size 1 x 1, followed by a batch normalization layer. A 2D maxpool layer having (2, 2) pool size is added. The maxpool layer decreases the spatial size of the subsequent feature map by accruing the generated features. The third convolutional layer has 32 kernels of size 3 x 3 with an additional batch normalization layer. Another 2D maxpool layer having (5, 5) pool size and stride value of 5 is added.

Next, flatten layers and a fully connected dense layer are added to accept output elements. This fully connected layer uses the ReLU activation function. The ReLU function is well expressed by (3).

\[ \text{ReLU}(x) = \max(0, x) \]  

The SoftMax classifier is the final layer in the classification block. The model used the ReLU activation function on every layer and the categorical cross-entropy cost function for loss optimization.

The use of ReLU does not initiate neurons having input values as negative. This approach makes the model more computationally efficient over tanh and sigmoid functions as it creates sparsity. Table 1 shows the layers, kernel size, and the number of neurons in the model.

### B. Model Optimization

The model optimization process is applied to make models more efficient and dependable to input data. The deep learning model that was created employed Stochastic Gradient Descent (SGD) as the optimizer for compiling the model. The SGD functions as a parameter in every training sample and delivers much faster performance. The SGD completes exclusive updates at a particular time.

The cross-entropy provides a better option for optimizing the cost function. For improved classification and prediction outcomes, the cross-entropy cost function is employed in the model.

| TABLE I: LAYER, KERNEL SIZE, AND NUMBER OF NEURONS |
|-----------------|-----------------|-----------------|
| Layer Name      | Kernel Size     | Number of Kernels |
| Conv2D          | 1 x 1           | 8               |
| BatchNorm       | -               | -               |
| Conv2D          | 1 x 1           | 16              |
| BatchNorm       | -               | -               |
| MaxPool2D       | 2 x 2           | -               |
| Conv2D          | 3 x 3           | 32              |
| BatchNorm       | -               | -               |
| MaxPool2D       | 5 x 5           | -               |
| Flatten         | -               | -               |
| Dense           | -               | 128             |
| Dropout (0.1)   | -               | -               |
| Dense           | -               | 26              |

### C. Model Training

In training the model, the He normal initialization [28] was applied to initialize every kernel towards achieving quicker convergence. The batch-normalization technique was applied to limit the convolution layers from learning less relevant features and avoid overfitting to the features.

Batch normalization (BN) is a technique to normalize activations amongst connecting layers of deep neural networks to enhance accuracy and training [29]. The addition of the batch-normalization layer aids in the reduction of covariance shifts within the network’s hidden kernels. The outcome from the previous layer is normalized via BN engaging the present batch statistics. This approach makes the weight distribution of the hidden kernel units or adapts to any data distribution variation. The BN layer also provides a level of regularization effect as weights are scaled depending on the batch mean and variance, which gives the weights stability. Fig. 4 shows the model training flowchart.

### VI. EXPERIMENT AND RESULT

The model is trained using the training and validation set representing 80% and 10% of the dataset distribution, respectively. The neural network completes classification tasks for the indigenous sign language in Nigeria by learning key features for individual classes in the dataset. The classification model attained excellent performance with no transfer learning. The downside to this technique is that if similar data exist for some classes in the subsequent classification task, the possibilities of data leakage will cause biases transfer learning.

In tackling this issue, the model was initialized via the He initialization for classification tasks. The model was trained for 30 epochs, with experimentation performed on multiple hyperparameters, including its regularization, and learning rate. The best values found for the initial learning rate are 0.0001. The model also employs a schedule learning rate which uses step decay to avert divergence and enhance the loss curve. The step decay takes in a learning rate and decreases by a factor of 0.1 after every 5th epoch.

Accuracy, confusion matrix, Recall, Precision, and F1 score were used in evaluating the model performance. Fig. 5 shows the plot of training accuracy against the validation accuracy. Fig. 6 shows the plot of training loss against the validation loss. Both plots were achieved using 30 epochs.
and FN. Table III shows the TP, FN, FP, and TN values obtained for Niger, Calabar, Enugu, Imo, Bayelsa, Bauchi, and Igbo.

For the multi-class classification model, observations from the accuracy plot showed that the graph attains more smoothness after every 5 epochs. This variation comes from the use of learning rate decay happening after successive 5th epoch. Decreasing by a 0.1 factor, the learning rate decay aids the model to converge optimally and stabilizes the learning process. The final model developed using the word dataset containing the indigenous sign language in Nigeria performed 95.67% training accuracy and 95.28% validation accuracy. Table II shows the evaluation outcomes.

A confusion matrix or error matrix for the classification model is shown in Fig. 7. The confusion matrix presents the class-wise performance of the developed model. The multi-class classification model was tested with a total test sample of 1500 images (10% of the word dataset). The 1500 samples comprise 15-word classes, each having 100 samples for the test.

From the confusion matrix in Fig. 7, it was observed that of the 1500 samples:

i. Test samples for Niger, Calabar, Enugu, Imo, Bayelsa, Bauchi, and Igbo were correctly predicted. The 100 samples predicted correctly for these words show its True Positive (TP) value as 100 and the False Negative (FN) value as 0. Calabar, Enugu, Imo, and Bauchi get a False Positive (FP) value of 0 as no sample was predicted incorrectly within its class. The True Negative (TN) value for the individual classes is calculated using the total sample (1500) minus the TP, FP,

For Ondo, Borno, Abia, Ekiti, Kastina, Kaduna, Anambra, and Are, the outcome for these classes shows varying True Positive (TP), False Positive (FP), and False Negative (FN). Ondo has the lowest TP value with 68. The highest misclass value of 31 was recorded between “Ondo” and “Are.” The values for the TP, FP, TN, and FN for class 1, 6, 7, 8, and 9 are shown in Table IV.

Based on the above observation, the following performance metrics were computed for the following class:

Ondo:

$$Precision (P) = \frac{TP}{(TP+FP)} = \frac{68}{68+4} = 0.94$$

$$Recall (R) = \frac{TP}{(TP+FN)} = \frac{68}{68+32} = 0.68$$

For the multi-class classification model, observations from the accuracy plot showed that the graph attains more smoothness after every 5 epochs. This variation comes from the use of learning rate decay happening after successive 5th epoch. Decreasing by a 0.1 factor, the learning rate decay aids the model to converge optimally and stabilizes the learning process. The final model developed using the word dataset containing the indigenous sign language in Nigeria performed 95.67% training accuracy and 95.28% validation accuracy. Table II shows the evaluation outcomes.

**TABLE II: EVALUATION OUTCOMES**

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Accuracy</td>
<td>95.67%</td>
</tr>
<tr>
<td>Validation Accuracy</td>
<td>95.28%</td>
</tr>
</tbody>
</table>

For the multi-class classification model, observations from the accuracy plot showed that the graph attains more smoothness after every 5 epochs. This variation comes from the use of learning rate decay happening after successive 5th epoch. Decreasing by a 0.1 factor, the learning rate decay aids the model to converge optimally and stabilizes the learning process. The final model developed using the word dataset containing the indigenous sign language in Nigeria performed 95.67% training accuracy and 95.28% validation accuracy. Table II shows the evaluation outcomes.

**TABLE III: TP, FN, FP, AND TN VALUES FOR NIGER, CALABAR, ENUGU, IMO, BAYELSA, BAUCHI, AND IGBO**

<table>
<thead>
<tr>
<th>Word</th>
<th>True Positive (TP)</th>
<th>False Positive (FP)</th>
<th>False Negative (FN)</th>
<th>True Negative (TN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger</td>
<td>100</td>
<td>3</td>
<td>0</td>
<td>1397</td>
</tr>
<tr>
<td>Calabar</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>Enugu</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>Imo</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>Bayelsa</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>Bauchi</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1400</td>
</tr>
<tr>
<td>Igbo</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>1398</td>
</tr>
</tbody>
</table>

**TABLE IV: TP, FN, TN, AND FN FOR ONDO, BORNO, ABA, EKITI, KASTINA, KADUNA, ARE, AND ANAMBRA**

<table>
<thead>
<tr>
<th>Word</th>
<th>True Positive (TP)</th>
<th>False Positive (FP)</th>
<th>False Negative (FN)</th>
<th>True Negative (TN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ondo</td>
<td>68</td>
<td>4</td>
<td>32</td>
<td>1396</td>
</tr>
<tr>
<td>Borno</td>
<td>82</td>
<td>0</td>
<td>18</td>
<td>1400</td>
</tr>
<tr>
<td>Abia</td>
<td>99</td>
<td>7</td>
<td>1</td>
<td>1393</td>
</tr>
<tr>
<td>Ekiti</td>
<td>98</td>
<td>2</td>
<td>2</td>
<td>1398</td>
</tr>
<tr>
<td>Kastina</td>
<td>99</td>
<td>20</td>
<td>1</td>
<td>1380</td>
</tr>
<tr>
<td>Kaduna</td>
<td>97</td>
<td>0</td>
<td>3</td>
<td>1400</td>
</tr>
<tr>
<td>Anambra</td>
<td>98</td>
<td>0</td>
<td>2</td>
<td>1400</td>
</tr>
<tr>
<td>Are</td>
<td>90</td>
<td>31</td>
<td>10</td>
<td>1369</td>
</tr>
</tbody>
</table>
The Precision, Recall, F1-score, and Support for all Word is shown in Table V.

<table>
<thead>
<tr>
<th>Word</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ondo</td>
<td>0.94</td>
<td>0.68</td>
<td>0.79</td>
<td>100</td>
</tr>
<tr>
<td>Niger</td>
<td>0.97</td>
<td>1.00</td>
<td>0.99</td>
<td>100</td>
</tr>
<tr>
<td>Calabar</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Borno</td>
<td>1.00</td>
<td>0.82</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Enugu</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Abia</td>
<td>0.93</td>
<td>0.99</td>
<td>0.96</td>
<td>100</td>
</tr>
<tr>
<td>Ekiti</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>100</td>
</tr>
<tr>
<td>Imo</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Kastina</td>
<td>0.83</td>
<td>0.99</td>
<td>0.90</td>
<td>100</td>
</tr>
<tr>
<td>Kaduna</td>
<td>1.00</td>
<td>0.97</td>
<td>0.98</td>
<td>100</td>
</tr>
<tr>
<td>Bayelsa</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Anambra</td>
<td>1.00</td>
<td>0.98</td>
<td>0.99</td>
<td>100</td>
</tr>
<tr>
<td>Bauchi</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>100</td>
</tr>
<tr>
<td>Are</td>
<td>0.74</td>
<td>0.90</td>
<td>0.81</td>
<td>100</td>
</tr>
<tr>
<td>Igbo</td>
<td>0.98</td>
<td>1.00</td>
<td>0.99</td>
<td>100</td>
</tr>
<tr>
<td>Micro Average</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>1500</td>
</tr>
<tr>
<td>Macro Average</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>1500</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>1500</td>
</tr>
</tbody>
</table>

The macro-averaged score is calculated by considering the total TP, total FN, and total FP of the model. The values of TP, FN, and FP for every class are added to evaluate the micro average score for Precision, Recall, and the F1 score.

\[
\text{Total } TP = (100 + 100 + 100 + 100 + 100 + 100 + 82 + 68) = 1431
\]

\[
\text{Total } FP = (3 + 2 + 4 + 7 + 2 + 20 + 31) = 69
\]

\[
\text{Total } FN = (32 + 18 + 1 + 2 + 1 + 3 + 2 + 10) = 69
\]

The micro-averaged score for Precision, recall, and F1-score is calculated as:

\[
\text{Micro Precision} = \frac{\text{Total } TP}{\text{Total } (TP+FP)} = \frac{1431}{1431+69} = 0.95
\]

\[
\text{Micro Recall} = \frac{\text{Total } TP}{\text{Total } (TP+FN)} = \frac{1431}{1431+69} = 0.95
\]

\[
\text{Micro F1} – \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{0.95 \times 0.95}{0.95+0.95} = 0.95
\]

The macro averaged score for the word model is evaluated to gives Micro Precision = 0.96, Macro Recall = 0.95, and Macro F1 – score = 0.95. The weighted-average score evaluated using the weighted mean of the measures gives the weighted – precision = 0.96, weighted – Recall = 0.95, and weighted F1 – score = 0.95.

Based on the performance metric values, the classification model for the word dataset has a precision of 0.96 (on a scale of 0 – 1), Recall of 0.95 (on a scale of 0 – 1), and an F1-score of 0.95 (on a scale of 0 – 1).

The F1 score was used because it emphasizes the false negatives and false positives in its evaluation process. Table 6 shows the average Precision, Recall, and F1 score of the developed model.

<table>
<thead>
<tr>
<th>Word Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

These experiments show the model considerably performed well with the quality input data and training process. The outcomes also validate the robustness and Precision in interpreting indigenous sign language in Nigeria using the proposed model.

VII. CONCLUSION

This paper presents an efficient real-time-based interpretation system for sign language of some indigenous words in Nigeria. A dataset containing word samples from the indigenous sign language in Nigeria was created to cater for training the model. This paper considered hand movement as the key feature, creating static image samples from continuous video stream subjected to image processing techniques. The model developed uses the convolutional neural network as the classifier. For the interpretation of signs, appearance-based and digital interpretation, the CNN proves to be a better choice. The proposed model recognizes and interprets individual signs as text, supporting the removal of any communication gap amongst signers and non-signers. The method developed has given excellent interpretation accuracies attaining a 95.67% training accuracy and 95.28% validation accuracy. The proposed interpretation system using the CNN classification approach achieves 0.95 of Precision, 0.95 of Recall, and 0.95 of F1-score (on a 0–1 scale). Towards improving the interpretation system, the next step is to gather more data for Nigeria's indigenous sign language. In consideration of approach and practicality, more focus will go into interpreting and recognition of idioms, phrases, and sentences available in the indigenous signs of Nigeria.

ACKNOWLEDGMENT

We acknowledge with deep gratitude the special class of the Redeemed Christian Church of God (RCCG) Kings Sanctuary, Akure, Nigeria, and the staff of the School for the Deaf, Akure, Nigeria, for their valuable assistance for this work.
REFERENCES


A. O. Olabanji obtained B. Eng in Electrical and Electronics Engineering from the Federal University of Technology, Akure, Nigeria in 2016. His research interest includes communication systems, artificial intelligence, digital signal processing, IoTs, and biomedical engineering. He is currently a postgraduate student in the department of Electrical and Electronics Engineering in the Federal University of Technology, Akure, Nigeria.

A. A. Ponnie obtained B. Eng. and M. Eng. Degrees in Electrical and Electronics Engineering from the Federal University of Technology, Akure, Nigeria in 1998 and 2003 respectively, and obtained PhD degree in 2011 from Tohoku University, Japan. His research interests include communication systems, digital signal processing, electronics, and biomedical engineering. He currently teaches in the department of Electrical and Electronics Engineering at both undergraduate and postgraduate levels in Federal University of Technology, Akure, Nigeria where he is currently an Associate Professor. He is a member of Nigeria Society of Engineers (NSE) and a registered engineer with the Council of Engineering Regulation in Nigeria (COREN).