Online Nigerian Languages Word-Level Character Recognition Using Deep Transfer Learning

Afeez Adekunle Soladoye[1,*,†], Ibrahim Adepoju Adeyanju[1,†], Bolaji Abijail Omodunbi[1,†] and Adedayo A. Sobowale[1,†]

[1] *Federal University Oye-Ekiti, Oye-Ekiti, Ekiti State. Nigeria*

Abstract

The application of artificial intelligence has cut across all areas of life birthing from the technological advancement and its wide acceptance and usage. The conversion of typed image text into machine readable format has become necessity as majority of typing and communication is being done via smartphones. Moreover the understanding of these text like translation warrant it being firstly recognized. As a result of that, many Nigerian languages are being used as the communication language or part of sentence during online chat or dissemination of information like tweets, LinkedIn and Facebook post, not everybody on these platforms are Nigerian and not all Nigerian on the platforms speaks the same language, so there is a need for recognition of the Language mixed within some other popular language Like English for easier understanding. This study developed three Nigerian Word character recognition using ResNet-50 an architecture of Convolutional Neural Networks. 100 different words per languages were acquired over the internet and converted to images using Python script. These images were preprocessing using normalization, transformation, rescaling, and feature extracted using Histogram of Oriented Gradient. The extracted feature was then fed into ResNet-50 for recognition and classification with an accuracy of 77% while Yoruba gave the best weighted precision of 89%. The evaluated result was further compared with some machine learning algorithms like K-nearest Neighbor and Support Vector Machine and the system using ResNet-50 outperformed the earlier mentioned model. This study gives a headway on the possibility for development of such multi-class recognition system. Future work should focus on using more words and more Nigerian Languages to expand the scope and improve the performance of the system.

Keywords

Nigerian Languages recognition, Optical Character Recognition, Deep Transfer Learning, Human machine communication

1. Introduction

The advancement of technology has brought about the need for conversion of typed or handwritten text images for machine usage through its readability, editability for manipulation. This brought about the concept known as Optical Character Recognition (OCR), this is an aspect

ICCBR AI Track'24: Special Track on AI for Socio-Ecological Welfare at ICCBR2024, July 1, 2024, Mérida, Mexico

*Corresponding author.
†These authors contributed equally.
afeez.soladoye@fuoye.edu.ng (A. A. Soladoye); ibrahim.adeyanju@fuoye.edu.ng (I. B. Adeyanju) bolaji.omodunbi@fuoye.edu.ng (B. A. Omodunbi); adedayo.sobowale@fuoye.edu.ng (A. A. Sobowale)
0000-0002-6349-5173 (A. A. Soladoye); 0000-0002-6163-2453 (I. B. Adeyanju); 0000-0003-1309-79X83 (B. A. Omodunbi)

of pattern recognition that helps in recognition of typed or handwritten text from digitalized image through scanning or image generation [1]

The Nigerians usually include their native words in the context of their post on different social media platforms, and inability to understand these native words might not enhance the understanding of the sentence and create communication barrier. Moreover, in order to preserve languages they have to be used in the social media spaces as that is where people of diverse origin and culture relate and interchange their cultural diversity for inclusion and adoption.

Africa is not behind as regards the technological advancement most especially in the application and research involving Artificial Intelligence. This field needs to be explored for the good of the people of Africa at large. The three Nigerian famous languages namely: Yoruba, Hausa and Igbo are not spoken in Nigeria alone but also in other west African countries like Benin republic, Ghana and other Countries like Somalia, Brazil, United Kingdom and United states of America among others. These languages are also well used for communication through different Social media platforms like Twitter, Facebook and LinkedIn. If these languages have this kind of recognition and acceptance, people with little understanding of these languages should not be deprived of their understanding, as they are natural languages spoken by different people of different origins.

Many researches have been conducted on Nigerian Language recognition using different conventional machine learning algorithms and deep learning algorithms. Adeyanju, Ojo and Omidiora [1] conducted a study on recognition system for English typewritten characters using Hidden Markov Models (HMM). Three sets of typewritten dataset were used. The system showed a recognition accuracy of 94.88%, 91.45% and 97.24% respectively for old memo, old war letter and newly typewritten essay. The recognition system recorded its best result at 0.8 thresholds. The system cannot recognize handwritten characters and formatted characters. Similar study was conducted by Ahmad *et al.* [2] for online character recognition, the system used words from two different databases that consist of 4,086 isolated digits, 10,685 isolated lower case letters, 10,679 isolated upper case letters and 410 EURO signs. It also contains 31,346 isolated words from a 197 word lexicon (French: 28,657 and English: 2,689). Neural network and SVM were evaluated on the database, SVM gave the best recognition accuracy on UNIPEN database dataset.

Adegunlehin, Asahiah and Onifade [3] presented in their research characterization of Yoruba named entity using conditional random fields. The study employed different features for recognition such as context- words, part of speech information. The study gave a weighted average precision of 89% when the part of speech information tag was not included and 81% when the surrounding words were discarded. This study shows that feature choice have a significant effect on the Yoruba name entity recognition as shown by the evaluation result. Oni and Asahiah [4] conducted a relevant research on the Yoruba words in printed text, this study used text from Bibeli Mimo a Yoruba version of the Bible. The dataset used consist of 4000 words for both training and validation, these words are types using three font styles namely: Times Roman, Ariel and Deja VuSans. Binarization, skew estimation and page segmentation were performed as the preprocessing techniques on the dataset. LSTM was used as the recognition model and the system gave a good character error rate across all the three font styles employed. Similarly, Ajao *et al* [5] carried out a research on recognition of handwritten Yoruba recognition, this study employed Hidden Markov Model as the recognition model for the handwritten characters. The study employed Yoruba words with diacritics and gave a good recognition rate.

Naseer *et al.* [6] proposed Balochi Non-Cursive isolated character recognition system using Deep Neural Network for Balochi script recognition for non-cursive characters. The system was compared with the baseline LeNet model and the results showed a precision of 96% and it trains rapidly compared to the baseline methods. Offline Yoruba word character recognition system was developed by Oladele *et al.* [7], this study used twenty-five different character per Yoruba alphabets simulated on MATLAB 2015a. Grayscale conversion, binarization, noise removal cropping and resizing, segmentation and skeletonization were the preprocessing techniques performed on the acquired character images. Zoning and gradient descendent were used for extortion of geometric features in the character, after which SVM was used as the classifier. The study gave a recognition accuracy ranging from 60-100%. Ajao, Okunade and Ajao [8] conducted a research that employed recurrent neural network for the recognition of handwritten Igbo character recognition. The study acquired these characters from student with a total of 3,600 scanned images (1,800 each for both upper and lower case letters). The images was converted to grayscale and resized to 50x50. The accuracy obtained from the system was not that good but gave a starting point for further study on recognition of igbo character.

From all the presented related works most of the works done on character recognition is either on English or Yoruba and majority of them did not used any deep transfer learning approach nor consider character recognition of multiple Nigerian Languages. This is quite important gap to fill as there is always usage of multiple language on the internet and if only types Yoruba words are being considered what would happen to the remaining languages where there are more than 250 spoken languages in Nigeria. This study aim to consider recognition of the three main Nigerian Languages using Optical Character Recognition techniques with deep transfer learning. This study contributes to the body of knowledge in the following ways:

i. Recognition of Nigerian language using Optical Character Recognition techniques
ii. Applying Deep transfer learning for Nigerian language word-level character recognition.
iii. Employing freshly acquired Nigerian words of Character recognition.
iv. Recognition of more than one Nigerian Language word-level character

This manuscript is organized as follows, after this introduction section tha provides brief insight about the study and relevant studies that have been done on the research areas, the methodology section follows, which discuss the stages involved in conducting this research, after which the result obtained using the deep transfer learning algorithm for word-level character recognition of 3 major Nigerian languages was discussed. Finally, the conclusion and recommendation ends the major sections of the manuscript after which acknowledgement and reference section follows.

2. Methodology

This study followed the optical character recognition system development pipeline of data acquisition, data preprocessing, features extraction and recognition. This is shown in Figure 1 for easier understanding and clarification.

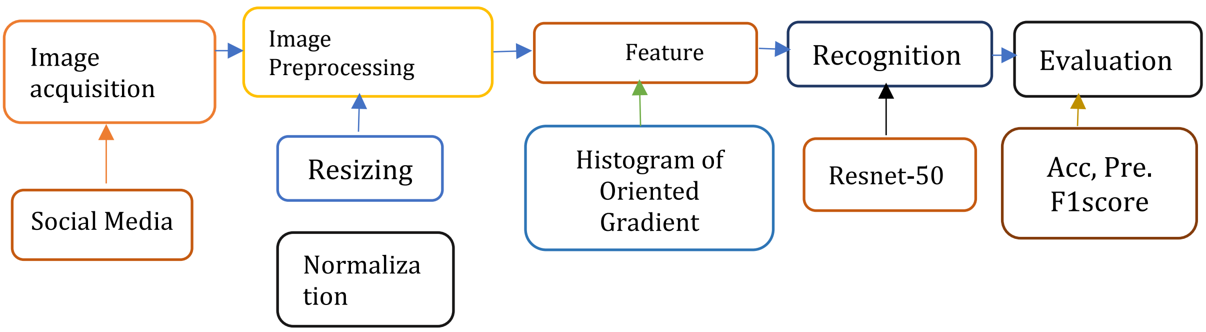


Figure 1: Overview of research methodology

2.1. Data Acquisition

This study randomly acquired 100 words each for Yoruba, Hausa and Igbo form different social media platforms like Facebook, Twitter and some words were randomly acquired online through different platforms by searching through Google among other. This approach of data acquisition was employed so as to ensure that only frequently used and publicly available words were captured so as to prevent bias and acquisition of unpopular words, this data acquisition stage was carried out by five different data collectors –student in the department- to ensure diversity, exploration of different platforms and thorough search. The acquired words were saved as word document for further preprocessing. Moreover, Arial black and Times New Roman font styles were used to types the words due to their bold nature and wide usage to enhance easier feature recognition and word classification. Some of the Yoruba words have diacritics while some Hausa and Igbo words also have character tone signs. Samples of words per the three main Nigerian are represented in Table 1.

Table 1: Sample of acquired word per Language

Class	**Image**	
Hausa	Tsangwama	Sa-in-sa
Igbo	biko	gini
Yoruba	ikarahun	ọrun

2.2. Image Preprocessing

The acquired words in textual format was firstly converted into images to enhance the use of Optical Character Recognition techniques. This text to image conversion was done through "word-as-image" python script. This script was initialized with font size of 12, width and height of 100 and 30 in pixels respectively, encoded with 'utf-8' with RGB colour orientation and saved in PNG format. The generated images were saved in different folders for processing and loaded into another file for implementation. The images were further preprocessed to enhance the character recognition and improve the performance of the model for language classification. The first preprocessing was grayscale conversion, this helped in converting the original RGB generated images to grayscale images for compression and reduce memory space consumed by RGB format. The images were further Normalized, resized and rescaled to ensure all the training and testing images are of the same dimension to prevent inconsistency. The images were converted into grayscale after loaded. Data augmentation was also performed on the preprocessed images which in turn increase the number of images to 400 per class, some of this transformations performed on the images are random horizontal flip, center crop and color jittering. Moreover, the images were further augmented so as to increase the volume of the available dataset for more accurate recognition.

2.3. Feature Selection

Feature selection is an important stage in image processing, as it helps in obtaining distinctive features from the images, it also serves as image dimensionality reduction by reducing the total dimension of the images to a smaller dimension with the removal and extraction of only the important and significant features from the whole image [9]. This helps in enhancing the performance of the recognition model and reduce computation time and resources. Histogram of Oriented Gradient is a feature extraction techniques used in computer vision for image processing which counts the occurrence of orientation of a gradient in the image's localized position. This technique was used as the feature selection techniques in this study to focus on the important and relevant features (the character) in the images and discard the irrelevant ones.

2.4. Recognition

Word-level character Recognition is the act of classifying and assigning classes/ labels to a specific set of images to determine whether they belong to one class or another based on the recognized characters in the images. This study is a multi-label recognition problem with the presence of three classes of words namely Yoruba, Hausa and Igbo. Image classification is process of learning similar features that individual word class possess and differentiate them from others. In order to achieve this, Residual Newtork-50 also known Resnet-50 a subtype of Convolutional Neural Networks architectures was used as the recognition model. ResNet-50 has 50-layers of CNN with 48 convolutional layers, 1 MaxPool layer and 1 average pool layer.

ResNet-50 architecture consists of the following element as its building blocks: a 7x7 kernel convolution along with 64 other kernels with a 2-sized stride, a max pooling layer with a 2-sized stride, 9 layers (with 3x3,64 kernel convolution, 1x164 kernels and 1x1256 kernels, these three layers are repeated 3 times making 9 layers), 12 layer (with 1x1128 kernels, 3x3128 kernels and 1x1512 kernels, these three layers were repeated 4 times), 18 layers (comprising of 1x1256 cores, 2 cores 3x3256 and 1x11024 cores iterated 6 times), another 9 layers (consist of 1x1512 cores, 33512 cores and 1x12048 cores iterated 3 times) an finally average pooling followed by fully connect layer with 1000 nodes that used softmax as its activation function . These layers are represented in ResNet-50 architectural diagram represented by Figure 2 for better understanding and clear representation.

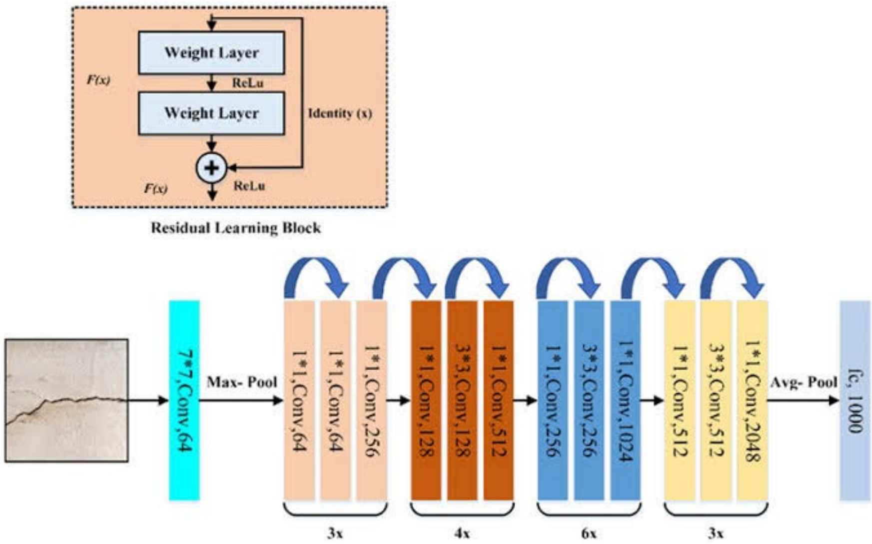


Figure 2: Architectural overview of ResNet-50

2.5. Experimental Setup

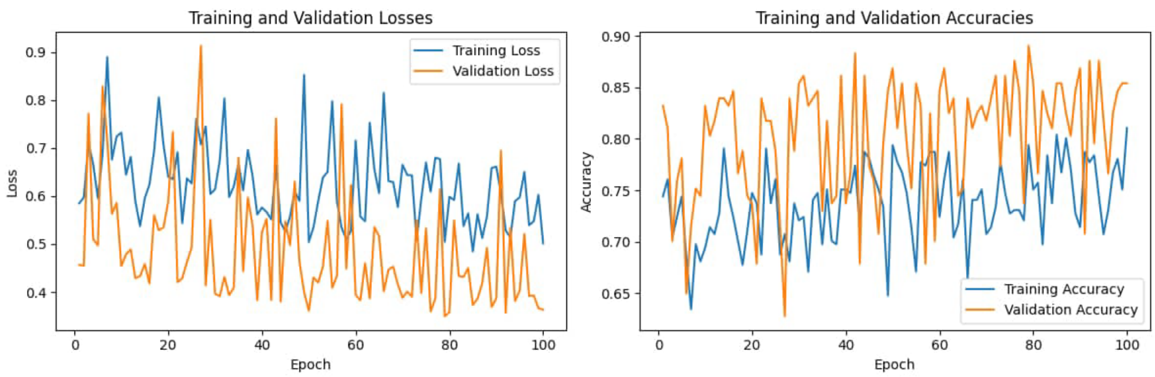
The study developed Nigerian Language recognition employed a total of 300 images with 100 per language with 20% of the data for testing. The system was implemented on Google Colab using Graphical Processing Unit (GPU) with Python v3.9 with importing of major libraries like Pandas, Skimage, Scikit-learn, ImageDraw used for images manipulation and preprocessing. The study was implemented on Intel core i5 @2.00GHz, 8GB RAM and Windows 10 Operating System.

3. Results and Discussion

The acquired data was implemented for classification using ResNet-50 as the classification model. The result obtained from this classification was presented in this section and further compare with other machine learning algorithms to compare their performance and show the model with the best classification performance.

As discussed earlier, ResNet-50 is an architecture of CNN with 50 layers and average pooling layer with the last layer being fully connected layer with 1000 nodes and softmax as the activation at this layer. This shows that the model is good for multi-label classification problem, which this study is an example.

Figures 3a and b represent the training and validation losses and accuracies of the classification model training ResNet-50 over 100 epochs.



(a) (b)

Figure 3: ResNet-50 Training and Validation Loss and Accuracy Graphs

As shown in Figure 3a the validation loss started around 0.47 while the starting point of that of training loss was above with its starting point being around 0.59. Around 30th epoch the validation loss approached 0.92 while the training loss also moved drastically around 5th epoch to 0.90. This fluctuation in their losses continues until the last epoch where the validation loss comes down below 0.3 and that of training loss was around 0.5. From this illustration, the training loss was higher than the validation loss and this implies that the system was well validated on the validations split.

Figure 3b also represents the training and validation accuracies across all the training epochs from 1 to 100. The validation accuracy started higher than that training accuracy, though around 26th epoch it goes down below 0.65 but it eventually when up toward 42nd epoch and maintained the high values till the last epoch. Same thing applicable to the training accuracy, however, the training accuracy wasn't that high as the validation accuracies but also finished above 0.80.

The results given by Figure 3 shows that the system was well trained on the dataset and the accuracy was still good even with the small size of the image dataset used for training the model, if more images of words per class was acquired the model would perform better and improve the fluctuation in the training accuracies and losses as observed in Figure 3.

The average evaluation metrics result obtained after the model was evaluated with the testing data is presented in Table 2 for clearer representation.

Table 2: Classification report of the Experimentation Evaluation

	Precision	Recall	F1-score	support
Hausa	0.73	0.77	0.75	100
Igbo	0.70	0.78	0.74	100
Yoruba	0.89	0.75	0.82	100
Accuracy			0.77	300
Micro avg	0.78	0.77	0.77	300
Weighted avg	0.78	0.77	0.77	300

Table 2 shows the classification report of the evaluated model on the images for words classification based on languages. The Yoruba class gave the best average precision of 89% followed by Hausa. The model average accuracy was 77% and the number of support for the whole testing split per language is 100 words, this increased due to the image transformation performed on both the training and testing split during the image preprocessing stage.

This performance obtained from evaluating the ResNet-50 model is quite impressive as implies that the system can be used for real-life language recognition which would make translation easier once recognized.

The performance of the model was further compared with other machine learning algorithms, for this comparison, this study designed a KNN model with varying values of k with the same preprocessed dataset used for the experimentation with ResNet-50, similarly Support Vector Machine was equally experimented with the dataset using Linear and Poly kernels for experimentation. This result obtained during the experimentation is presented in Table 3.

Table 3: Comparison with other Machine Learning Models

S/N	Model	Accuracy (%)
1	k-Nearest Neighbor	57
2	Support Vector Machine	51
3	Developed model (ResNet-50)	77

From the result presented in Table 3, it shows that the developed word-level language character recognition system that employed ResNet-50 as the classification model outperformed the other models compared with and the accuracy difference is much which showcase that impressive performance of the ResNet-50 for the recognition of different Nigerian languages.

4. Conclusion and Future work

The use of Nigerian Languages is now widely employed for conversation and communication over the Internet using different platforms. However, many Nigerian could not even distinguish between two related Nigerian Languages and inability to recognize or distinguish a language from another would birth inability to understand or being able to translate this. This study developed word level language recognition using deep transfer learning for classification of these words. With the text to image python script used in this project, words from different languages acquired over the internet were converted to images, preprocessed and used HOG for features extraction before it was classified using ResNet-50 as the deep learning transfer algorithm. Recognition of Nigerian Languages proposed in this study would help in easy recognition of three main Nigerian languages on the internet and enhance understanding of the reader. Future works should focus on acquiring more words per language to increase the dataset and ensure that Yoruba words are written with diacritics as that gives it a distinction among other languages. More deep learning algorithms should also be used for recognition and compare their performance to know the best algorithms with the best recognition and classification accuracy.

References

[1] I. A. Adeyanju, S. O. Ojo and E. O. Omidiora. "Recognition of typewritten characters using Hidden Markov Models". British Journal of Mathematics and Computer Science, 12.4 (2016): 1-9

[2] A. R. Ahmad, C. Viard-Gaudin, M. Khalid and E. Poisson. "Comparison of Support Vector Machine And Neural Network in Character Level Discriminant Training for Online Word Recognition". Accessed on 22nd February, 2024 from https://www.researchgate.net/publication/228544717

[3] E. A. Adegunlehin, F. O. Asahiah and M. T. Onifade. "Investigation of Feature Characteristics for Yoruba Named Entity Recognition System". In the Proceedings of 2019 AICTTRA Conference, Nigeria (2019)

[4] O. J. Oni and F. O. Asahiah. "Computational Modelling of an Optical Character Recognition System for Yor`ub´a Printed Text Images". Scientific Africa. (2020)

[5] J. F. Ajao, S. O. Olabiyisi, E. O. Omidiora and O. O. Okediran. "Hidden Markov Model Approach for Offline Yoruba Handwritten Word Recognition" British Journal of Mathematics & Computer Science 18.6 (2016): 1-20

[6] G. J. Naseer, A. Basit, I. Ali and A. Iqbal. "Balochi Non-Cursive isolated character recognition using Deep Neural Network" International Journal of Advanced Computer Science and Applications, 11.4 (2020): 717-722

[7] M. O. Oladele, T. M. Adepoju, O. M. Olatoke and O. A. OjO. "Offline Yorùbá Handwritten Word Recognition Using Geometric Feature Extraction and Support Vector Machine Classifier". Malaysian Journal of Computing, 5.2 (2020): 504-514,

[8] J. F. Ajao, O. A. Okunade and A. O. Ajao. "Recurrent Neural Network (RNN) for Igbo Handwritten Character Recognition". The Pacific Journal of Science and Technology, 24.2 (2023)

[9] O. O. Bello. "Development of a canny edge and histogram of oriented gradient based sign language recognition system". Master dissertation, Federal University Oye-Ekiti, Computer Engineering (2021)