

Development of an Ethnic Classification Model Based on Traditional Attires

Segun Aina

Department of Computer Science &
Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria
s.aina@oauife.edu.ng

Aderonke Lawal

Department of Computer Science &
Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria
arlawal@oauife.edu.ng

Samuel Okegbile

Department of Computer Science &
Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria
okesam70@yahoo.com

David Asamu

Department of Computer Science &
Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria
davash001@gmail.com

Adeniran Oluwaranti

Department of Computer Science &
Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria
niranoluwaranti@oauife.edu.ng

Abstract— Ethnicity classification involves a comparison process of subjects belonging to a certain ethnic group with the subjects of another ethnic group. Image-based classification has been a popular approach to ethnicity cataloguing, with a focus on facial features, since they convey the most direct impression of an individual. However, in cases where the faces are not usually well captured in an image, such an approach makes it difficult to correctly predict the ethnicity of the subject in the image. This study, therefore, approaches the problem of ethnic classification by exploring traditional attire as a marker for cultural identity. A convolutional neural network-based model was developed to identify traditional Yoruba attires in images. The model was trained using 1000 clothing images belonging to two classes. The outcome of the proposed approach showed an estimated accuracy of 82percent and a precision of 0.837.

Keywords—Image Classification, Ethnicity, Nigerian Traditional Attire, Convolutional Neural Network

I. INTRODUCTION

According to [1], understanding the cultural and ethnic diversity that of people and their various worldviews is essential to developing a high level of international mindedness. An ethnic group or ethnicity is a set of people who identify with each other, usually on the basis of a presumed common genealogy or ancestry or on similarities such as common language or dialect, history, society, culture or nation [2].

With a current population that is slightly above 200 million, Nigeria is the most populous country in Africa and it is composed of multiple ethnic groups and cultures. Nigeria has about 250 ethnic groups and languages. Of these, the Igbo, Yoruba, and Hausa tribes are dominant in the country. These ethnic groups have unique traditional styled-clothing, which distinctively identifies them [3].

Clothing plays a vital role in the self-realization of every person. According to [4], clothing is a means of covering and protection, and it serves the role of communicating the social status, lifestyle, or membership of a particular culture or subgroup (ethnicity).

In Nigeria, the western style of dressing has become popular on a day-to-day basis. However, traditional

attires are worn for special occasions like festivals and weddings. On such occasions, the traditional attire serves the role of cultural identity, showing how the individual perceives himself or herself with respect to his or her ethnic group, nationality, religion and social status. This cultural identity cuts across several interconnected facets of life.

This study focused on the traditional attires of the Yoruba Ethnic group. It is one of the largest ethnic groups in Africa, found predominantly in southwestern Nigeria, other parts of Africa and some parts of Brazil.

In Yoruba culture, traditional clothes are often referred to as *Aso Ibile*. These are of two major types; Kijipa and Aso-Oke (Aso Ofi). Aso-Oke is worn on special occasions, while Kijipa is worn when engaging in menial jobs. Although Aso Ofi is of different types, the three most popular ones are Alaari, Sanyan and Etu. The three types can easily be differentiated by colour. Etu is dark blue with stripes, Sanyan is carton brown with white stripes and Alaari is crimson. Yoruba men wear Buba, Esiki with Dandogo, Agbada, Gbariye, Sulia, and Oyala. Some of their Sokoto are called Lambe, Gbanu, Sooro. Their caps include Gobu, Tinko, Abeti-Aja, Oribi, Bentigoo. Fig. 1 shows a Yoruba man dressed in the Aso-Oke traditional attire, wearing Abeti-Aja cap.

The most commonly found dresses among women are Iro and Buba or wrapper with blouse. They also have Gele and Ipele to complete their dressing. Apart from these, they have Iborun or Ipele according to [5] as shown in Fig. 2.

This study utilized image classification which involves identifying the output class of ethnicity to which an image belongs, or alternatively, the probability that the input image is of a particular ethnic (class) or otherwise. Image-based Ethnicity Classification can be regarded as a computer vision task. [6] described Image classification as the process of cataloguing images into one of a number of predefined classes. Convolutional Neural Network (CNN) involves deep learning algorithms that have become highly desired for computer vision-related tasks. CNN focuses on multi-scale and high-level image depiction that has an edge over the orthodox image processing algorithms which emphasizes low-level features in images.



Fig. 1: Yoruba man in traditional attire (Aso-Oke with Fila Abeti-Aja)

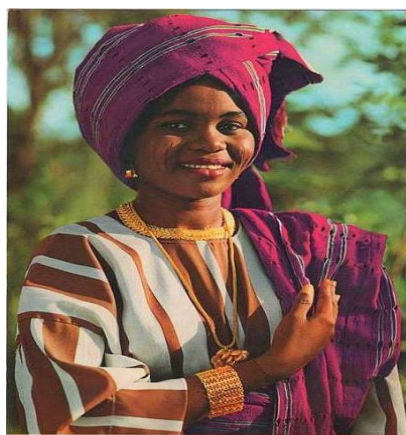


Fig. 2: Yoruba Woman in the traditional attire of Iro and Buba (Aso-Oke) with Gele and Ipele.

II. RELATED WORKS

Many classification techniques have been developed for image classification. Some of the most used techniques across literature are Support Vector Machine (SVM), Bag of Features, Artificial Neural Network (ANN), Convolutional Neural Network, Fuzzy Classification, and Decision Tree (DT).

In [7], apparels were classified using Convolutional Neural Networks, starting with an image containing clothing items while presenting a list of similar clothing

items available to buy online. The proposed model was able to correctly classify 41.1% of the images in the test set. The classification was further divided into style classification and attribute classification with the aim of improving the accuracy of the model. [8] also utilized Convolutional Neural Network for Apparel classification which was achieved by retraining the final layer of GoogLeNet for the task. The accuracy obtained was 98.5% for training the model and 98.22% for the testing model.

In a similar work carried out by [9], a framework to tackle the problem of clothing style classification in natural scenarios was proposed.

A combination of super-pixel segmentation, saliency detection, and the GMM model was introduced to reduce the interference of the background while working on clothing style recognition. An average accuracy of 70% was obtained for the 5 clothing styles considered.

The performance of the bag of features approach against the deep learning approach was compared in [10] using the Convolutional Neural Networks. The study showed that Convolutional Neural Network performs better in a machine learning framework as compared to a Bag of Features when extracting features in an image for classification.

III. METHODOLOGY

The approach used for Ethnicity Classification in this work was based on the classification of respective ethnic attires and as such can be regarded as an image classification problem. A Convolutional neural network (CNN) was trained for the purpose of the classification.

A. Data Pre-Processing

Data (images) from the database were made to undergo some transformations performed on the dataset before they can be used for training. This is done with the aim of increasing the performance of the machine learning model and speeding up the training process. The pre-processing techniques that were carried out sequentially on the image data are discussed in the sub-sections that follow.

In [11], the general architecture of a Convolutional Neural Network as shown in Fig. 3 was presented. The Convolutional Neural Networks were made up of multiple building blocks. These include the convolution layers, pooling layers, and fully connected layers. A Convolutional neural network is designed to learn spatial hierarchies of features automatically and adaptively by employing backpropagation.

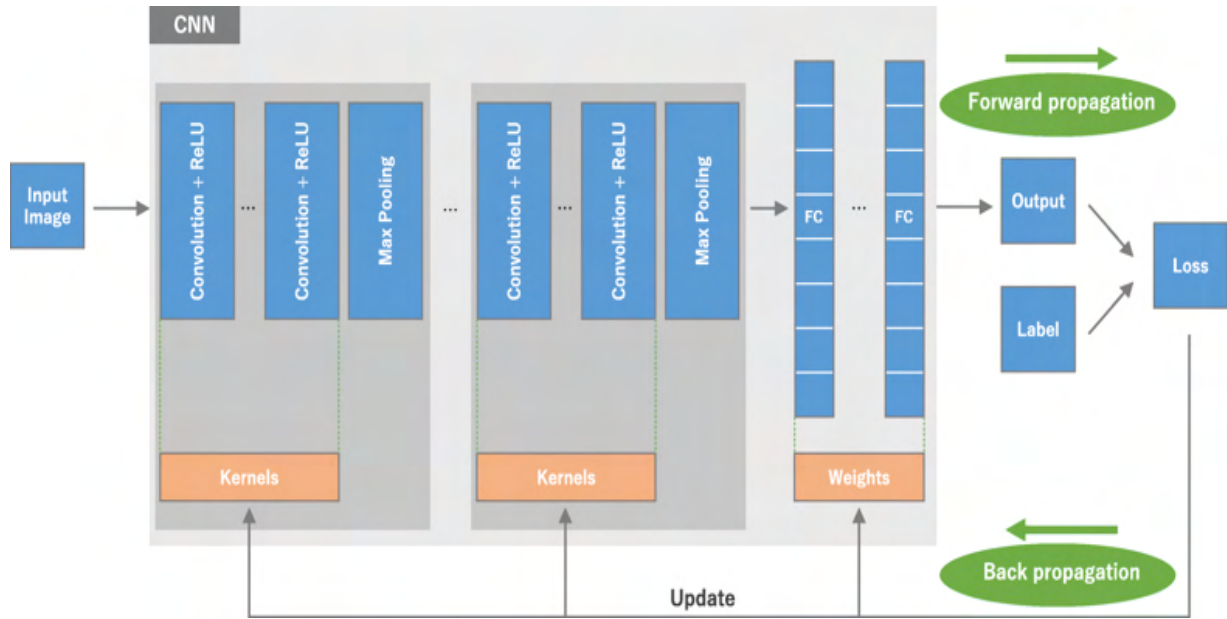


Fig. 3: An overview of the architecture for a convolutional neural network (CNN) [11]

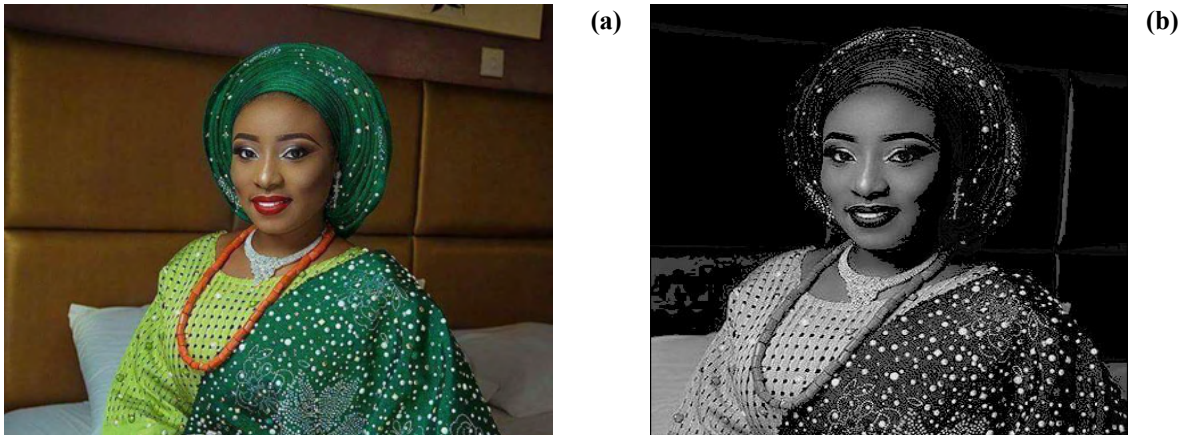


Fig. 4: (a) Original input image (b) Resulting image after Localization and Cropping

1. Localization of Clothing Region and Image Cropping

The purpose of the localization step was to remove unrelated information such as the background in the picture. Localization helps to identify and extract the clothing itself. The performance of an image classifier could significantly be influenced by the presence of noise in the form of natural backgrounds and fashion models in the training dataset. Fig. 4(a) shows the original image before the localization step, while fig. 4(b) shows the resulting image after localization and cropping was applied.

2. Data Augmentation

Data Augmentation therefore involved the process of generating samples by transforming the training data, with the goal of refining the precision and robustness of classifiers. The existing data-set was augmented with perturbed types of current images, including scaling and rotating.

In this study, a python library named Augmentor was used for the augmentation of the 1000 images collected during the study. The data Augmentor ensured that the neural network was exposed to the wide variety of images during training. This helped the neural network to identify the important characteristics for each of the classes in the training data-set.

The deep neural network approach to classification was used for the ethnic classification considering the size of the training data available in this study because the performance of the classification

model is closely related to the size of the training data. The same number of images in the dataset was the same before and after augmentation.

3. Image Resizing and Normalization

The input images in the dataset were resized while ensuring that the important information in the images was retained. This was done to speed up the training process.

The higher the quality of the training images, the more neural network nodes, and computing power is required. As such, the sizes of the images in the dataset were reduced before training.

Normalization ensures uniform data distribution for all input parameters (pixels in this case). This was done to ensure that the training converge faster. The normalized pixel value (z), was obtained by subtracting the mean value (μ) from each pixel to be normalized (x) and dividing the outcome by the standard deviation (σ). The normalization process is depicted in Equation (1) as:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

4. Data Splitting

It is important that a machine learning model is able to generalize accurately because when deployed, the model developed would be fed with data it has not previously encountered. As such, the extracted knowledge during training has to be generic enough. Hence, this study took note to avoid poor generalization that is often characterized by over-training. This was prevented by splitting the dataset. For the purpose of this study, the dataset was split into 3 sets.

- i. Training Dataset: This is the sample dataset used to train the model. The model is fitted by adjusting the weights and biases of the convolutional neural network.
- ii. Validation Dataset: The sample of data that was used to improve the performance of the model while tuning the model hyperparameters.
- iii. Test Dataset: This sample of the data was used to provide an unbiased evaluation of a final model. It was used to evaluate the performance of the model on images that had never been fed into the model previously. This was done after the model was completely trained using the train and validation sets.

In this study, 80% of the dataset was randomly selected for the training and validation of the model, while the remaining 20% was used to test the model.

B. Model Training and Validation

A Convolutional Neural Network (ConvNet/CNN) was trained for the purpose of the classification. Each input image was fed to the Deep Learning algorithm which then assigned weights, filters, and biases to various aspects of

the image with the goal of differentiation. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

As shown in Fig. 5, the following describes the high-level steps used to build the CNN model:

- i. Create convolutional layers by applying kernel or feature maps.
- ii. Apply Max pooling for translational invariance.
- iii. Flatten the inputs.
- iv. Create a fully connected neural network.
- v. Train the model.
- vi. Predict the output.

The convolution layer is responsible for feature extraction from the input image. This was achieved by performing a series of mathematical transformations (convolution, followed by activation). At the pooling stage, the convolved feature is further down-sampled by the convolutional neural network to speed up processing. Information critical in the features is preserved but the dimensions of the feature map are reduced. The pooling process makes use of the Max Pooling algorithm. The convolutional neural network ends with fully connected layers. Two layers in a CNN are “fully connected” if every node in the first layer is connected to every node in the second layer. The fully connected layers perform the classification task. The classification is based on extracted features during convolution. The final prediction is a value between 0 and 1 indicating the probability for each of the classification labels.

This is achieved by adding a softmax activation function to the final fully connected layer.

Validation was carried out to tune the hyper-parameters used in the training process of the model until the optimal performance was reached. The model was validated by using the K-Fold Cross Validation technique.

As shown in Fig. 6, the k-fold technique did not require a separate validation dataset, but instead at each iteration, it selects k - items from the training were used for validation. This approach was favored because of the limited size of the dataset. Since the images that would have used to create a validation set can now be added to the training set.

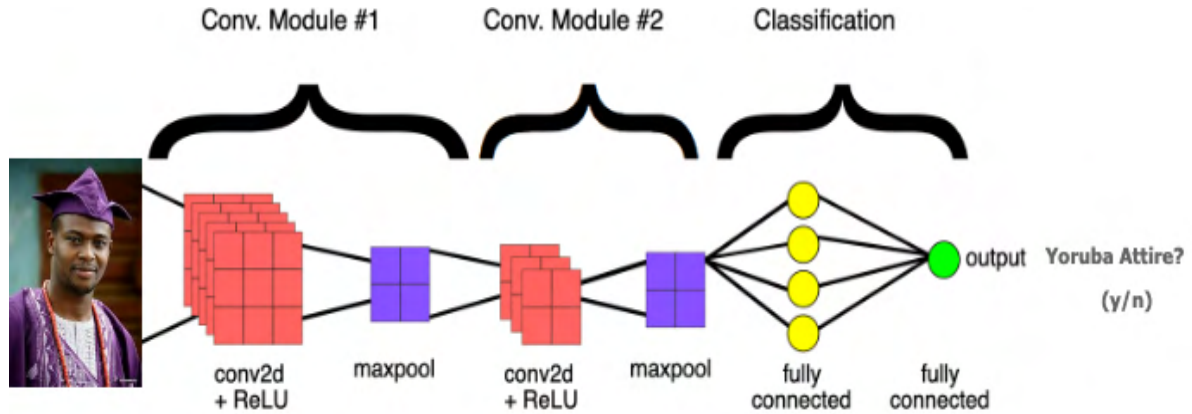
IV IMPLEMENTATION AND RESULTS

The implementation of the ethnic classification system involves steps for data collection, training, and testing as well as evaluation.

A. Data Collection

Images were collected by querying the Google Search API with keywords matching the description of traditional Nigerian attires. Some of the keywords used include:

manually filtered for outliers. A total of 1000 images belonging to two classes (Yoruba & Non-Yoruba) were collected. Fig. 7 shows a cross-section of some of the images that make up the dataset.



“Traditional Nigerian Attires”, “Yoruba Agbada”, “Igbo Wedding”, “Iro and Buba”, etc. The collected images were

Fig. 5: Architecture of Convolutional Neural Network

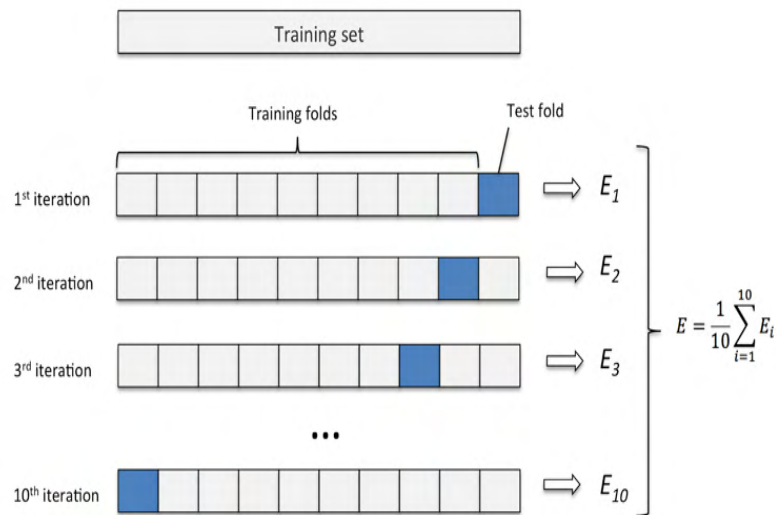


Fig. 6: Diagram of k-fold cross-validation with k = 10, [12].

B. Training and Testing

In this study, Keras library was used in developing the Convolutional Neural Network Model. Keras is a powerful and easy-to-use free open-source Python library for defining, building, and training deep Neural Network models.

The steps involved include loading the dataset images; and defining/compiling the Keras Model. This is followed

by training and fitting the model. Lastly, evaluation of the model and making predictions follow.

At the model-training phase, a model checkpoint and early stopping method were used to monitor the training at each parameter iteration. This promptly put a halt to the model once it was not improving in accuracy.

The following hyperparameters were considered and fine-tuned during the course of the training:

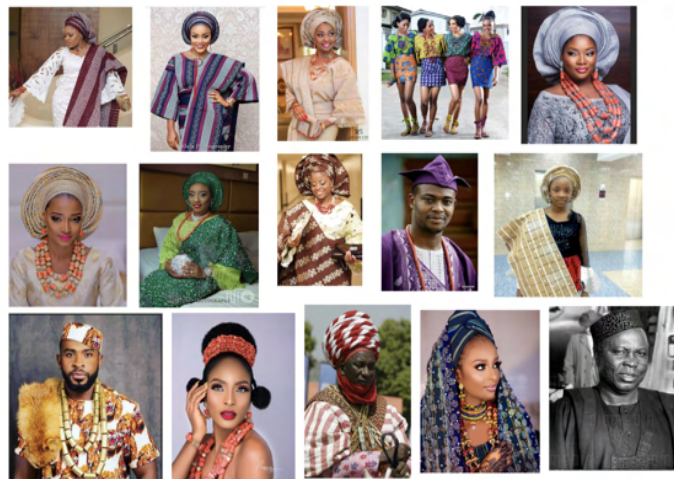
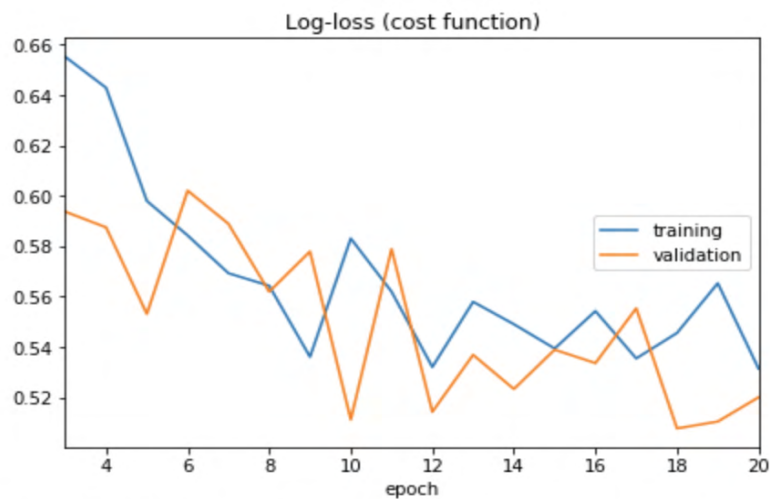


Fig. 7: Sample Images from Dataset

- i. The number of dense layers
- ii. The number of convolution layer
- iii. The number of nodes in each layer
- iv. The number of training epochs
- v. The batch size
- vi. The image size

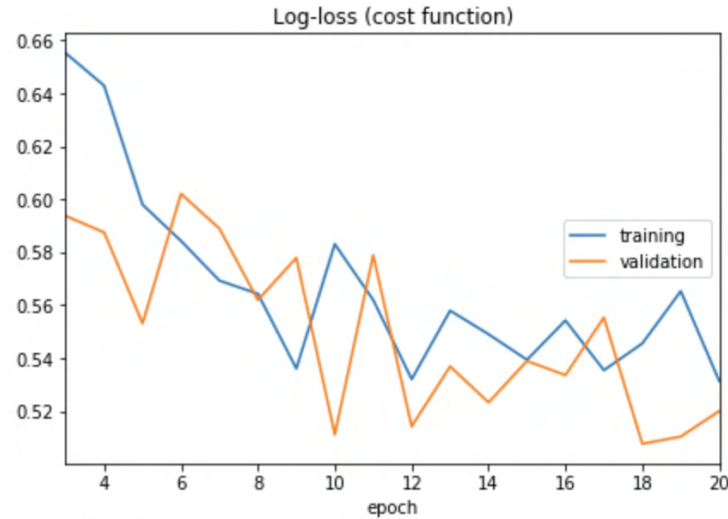
Fig. 8 below shows a graph of the cost function plotted against the number of epochs. This shows the log-loss

reduction as the number of epochs increases during the course of the training. A minimum log-loss value of 0.531 was recorded during the training step and the minimum value of 0.508 was recorded for the validation step. Fig. 9 shows a graph of the accuracy of the model evaluated at the end of each training epoch. A maximum validation accuracy of 0.819 was recorded during the training process.



Log-loss (cost function):
 training (min: 0.531, max: 0.694, cur: 0.531)
 validation (min: 0.508, max: 0.685, cur: 0.520)

Fig. 8: Snapshot showing model log-loss during the training process



Log-loss (cost function):
training (min: 0.531, max: 0.694, cur: 0.531)
validation (min: 0.508, max: 0.685, cur: 0.520)

Fig. 9: Snapshot showing model accuracy during the training process

VE EVALUATION

The performance of the classification model developed was evaluated using precision and accuracy metrics. This is summarized using the confusion matrix shown in Fig. 10. As seen, we have 430 True Positives (TP, predicted positive and are actually positive) and 389 True Negatives (TN, predicted negative and are actually negative). These can be used to calculate the accuracy and precision of the model as shown in (2) and (3) respectively. The accuracy of the model is a good indicator of performance because the classes are balanced i.e they have equal importance and probability of occurrence. Precision gives the percentage of

positive instances out of the total predicted positive instances. It is a way to find out "how much the model is right when it says it is right".

The model accuracy is obtained as:

$$\frac{TP + TN}{TP + FP + TN + FN} = \frac{430 + 389}{430 + 84 + 389 + 97} = 81.9\% \quad (2)$$

Similarly, the model precision is obtained as:

$$\frac{TP}{TP + FP} = \frac{430}{430 + 84} = 83.7\% \quad (3)$$

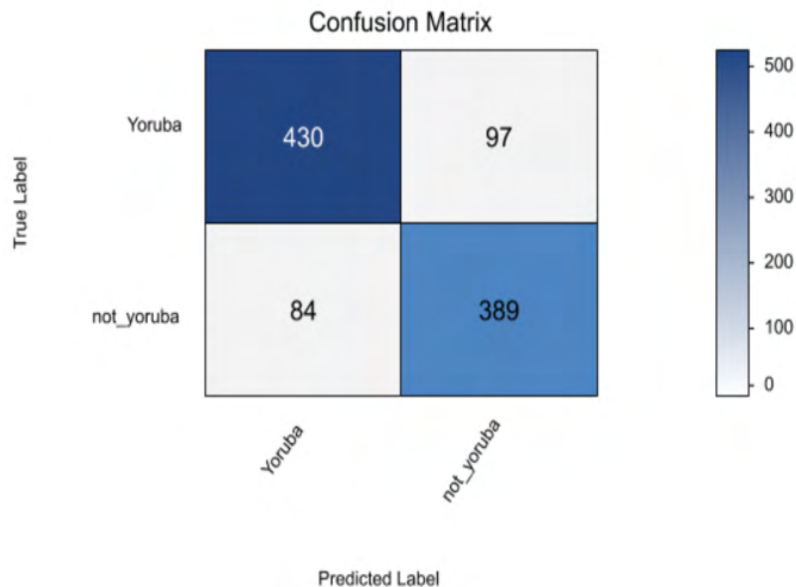


Fig. 10: Confusion Matrix of Classification Model

V CONCLUSION

The result obtained from the evaluation done shows that the classification model was able to distinguish different ethnic groups based on their traditional attires with an accuracy of 81.9%. The study, therefore, suggests that traditional attire can be considered not only as a marker for cultural identity but can also be used for ethnicity classification with a reliable level of accuracy.

In the classification of Yoruba and Non-Yoruba traditional Nigerian attires, it was discovered that grayscale preprocessing of the input images improves the accuracy of the result. Grayscale preprocessing involves making the colors in images to contain only grey shades ranging from pure white on the lightest end to pure black on the opposite end (achromatic). Grayscale images contain only shades of gray and no color. This improvement in accuracy can be attributed to the fact that color is not a distinguishing factor among the traditional attires of the tribes.

Future works include extending the model to support the two other major ethnic groups in Nigeria (Igbo and Hausa) and deploying the model as part of a multi-modal system for ethnicity classification in Nigeria.

REFERENCES

- [1] I. Hill, "Evolution of education for international-mindedness," *Journal of Research in International Education*, vol. 11 no. 3, pp. 245 – 261, 2012.
- [2] J. Peoples, and G. Bailey, *Humanity: An Introduction to Cultural Anthropology*. Wadsworth Cengage Learning, 2014.
- [3] T. Falola, *Culture and Customs of Nigeria*. Greenwood Publishing Group, 2001.
- [4] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. "Apparel Classification with Style," *Computer Vision- ACCV 2012: 11th Asian Conference on Computer Vision*, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part IV vol. 11, pp. 321-335. Springer Berlin Heidelberg, 2013.
- [5] A. O. Ogunsanya, "Changing Patterns in Feeding, Dressing and Naming among Yoruba of South-Western Nigeria since 1960," *Nigerian Journal of Social Studies*, vol. 18, no. 2, 2015.
- [6] P. Kamavisdar, S. Saluja, and S. Agrawal, "A survey on image classification approaches and techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 1, pp. 1005-1009, 2013.
- [7] R. Patki, and S. Suresha, "Apparel Classification using CNNs," unpublished results, 2016.
- [8] S. G. Eshwar, G. P. Gautham, A. V. Rishikesh, N. A. Charan, and V. Umadevi, "Apparel Classification Using Convolutional Neural Networks," *International Conference on ICT in Business Industry and Government*, IEEE conference, pp. 1 – 5, 2016.
- [9] L. Sun, X. Wu, H. Chen, and Q. Peng, "Clothing style recognition using fashion attribute detection," *EAI Endorsed Transactions on Ambient Systems*, vol. 2, no. 5, pp. 145 – 148, 2015.
- [10] S. Loussaief, A. Abdelkrim, "Deep learning vs. bag of features in machine learning for image classification," *International Conference on Advanced Systems and Electric Technologies*, IEEE, pp. 6-10, 2018.
- [11] P. Lanillos, D. Oliva, A. Philippsen, Y. Yamashita, Y. Nagai, and G. Cheng, "A review on neural network models of schizophrenia and autism spectrum disorder," *Neural Networks*, vol. 122, pp. 338–363, 2020.
- [12] J. M. Ashfaq, and A. Iqbal, "Introduction to Support Vector Machines and Kernel Methods", publication at <https://www.researchgate.net/publication/332370436>, 2019. Retrieved on April 28, 2020.