INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Mobile Application Deployment of CNN-based Food Classification System for Visually Impaired Individuals: A Real-Life Application for Nigerian Swallow Foods

OLASINA, Jamiu Rotimi[1], ALIU Olaniyi Habib[2], OLAIYA Olayinka Oluwaseun[3]

[1,2,3]Department of Computer Engineering, Federal Polytechnic, Ilaro, now Federal University of Technology, Ilaro, Ogun State, Nigeria.

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150600117

Received: 28 June 2026; Accepted: 03 July 2026; Published: 16 July 2026

ABSTRACT

The study addresses the significant challenge of food identification faced by visually impaired individuals, particularly in culturally specific contexts like Nigerian cuisine. A convolutional neural network (CNN) classification system was developed to categorize Nigerian swallow foods and assess their freshness (Fresh, 4 Hours Old, 1 Day Old) across 60 fine-grained classes using a dataset of 22,823 items. Six light CNN architectures were trained and evaluated based on metrics such as accuracy, macro-averaged precision, recall, F1 score, and AUC-ROC, alongside their parameter count and inference latency for deployment suitability. Among these, MobileNetV2 emerged as the top performer, achieving a test accuracy of 99.78%, a macro F1 score of 0.9979, and a perfect AUC-ROC of 1.0000, while maintaining a moderate parameter count of 2.33 million, making it suitable for mobile applications. LeNet-5 and ShuffleNetV2 also demonstrated strong performances with accuracies above 98.9%. In contrast, CCFNN exhibited poor performance with only 23.54% accuracy due to capacity issues. The best-performing model, MobileNetV2, was integrated into an application named Naija Food Eye, which facilitates food identification and freshness assessment for the visually impaired, providing confidence scores, top-5 predictions, and quick response times. The findings confirm that lightweight CNN models can effectively assist visually impaired individuals in recognizing foods within a culturally rich culinary landscape, offering a viable, deployable solution.

Keywords — Assistive Technology, Convolutional Neural Network, Deep Learning, Food Classification, Mobile Application, Nigerian Swallow Foods, Transfer Learning, Visually Impaired.

INTRODUCTION

Food identification is easy for people who can see, but for people with visual impairment, including 39 million who are completely blind, it can be a source of dependency and loss of autonomy in nutrition [1]. This is made harder in sub-Saharan Africa because there are few assistive technologies that are designed for indigenous food cultures. One of the biggest populations of the visually impaired in Africa is in Nigeria. Many starchy foods made of either cassava, yam, plantain, or wheat flour, such as Eba, Amala, Fufu, Semo, and Pounded Yam, are the staple foods of Nigeria but can be difficult to distinguish even for a sighted person without tactile or verbal cues. The Convolutional Neural Networks (CNNs) have given a new way to recognize images and are constantly outperforming hand-crafted features to classify food items [2] and [3]. With the increasing penetration of smartphone devices in Nigeria, there is a ready platform to deploy assistive solutions based on CNN with real-time, audio-based dietary feedback. However, current food classification studies mostly concentrate on the food cultures of the Western, East Asian, and South Asian cuisines [4, 5, 6] with little attention paid to the food cultures of Nigeria and Africa. This gap leaves the visually challenged Nigerians to rely on the sighted ones to identify food. This study tries to fill this gap by creating an application using CNN that is able to classify the swallowed foods found in Nigeria and provide audio feedback to the visually challenged.

Calorie counting systems [7] for global food, nutritional assessment systems [8], and dietary monitoring systems [9] are trained on external data sets, which are of no relevance to diets in Nigeria, which focus on the "swallow"

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

foods. Food identification can also not be solved by existing ATs for the visually impaired Nigerians, such as white canes, Braille, and screen readers. This deficiency is important because (i) Nigeria has a large population of visually impaired people due to preventable causes such as cataracts, glaucoma, onchocerciasis, etc.; (ii) foods in different classes are all similar in appearance, requiring special models; and (iii) it is expected that mostly people will adopt smartphones as a deployment tool.

The purpose of this research is to design, develop, and deploy a web based application that will be able to classify the swallow food in Nigeria using CNN. Objectives: (i) to compile and pre-process an image data set of at least five major food classes that are swallowed; (ii) to design and train a CNN model with transfer learning optimized for the task; (iii) to test the model performance using web application; (iv) to assess the model performance using accuracy, precision, recall, and F1 score.

The results of this research present a cheap and culturally appropriate solution to nutritional autonomy and the first CNN-based dataset and web application for Nigerian swallow foods. In addition, it follows the UN CRPD and Nigeria's Discrimination Against Persons with Disabilities Act 2018, as well as the SDGs 3 and 10 and the progress of regional food classification [4, 5], transfer learning with low-resource datasets [10, 11], and web application deployment for deep learning [7, 12].

LITERATURE REVIEW

The fields of application for food classification include dietary monitoring, dietary assessment, calorie estimation, food safety, and assistive technology. In early systems, handcrafted feature extraction like SIFT, HOG, Gabor filters, and color histograms were used along with classifiers like SVM, MKL, or KNN; however, the high intraclass variance and low interclass variance of food images were a problem [2]. CNNs revolutionized the accuracy of the classification task; Memiş et al. [2] reported that ResNext-50 with ImageNet pre-trained weights achieved 87.7% accuracy on UEC Food-100, which is significantly better than the traditional classification methods and points to the importance of transfer learning in small data sets. Sultana et al. [12] surveyed the methods used to estimate food values from images, highlighting that CNNs and architectures built upon transformer models are the two main approaches and revealing that the non-Western food varieties are underrepresented in the available datasets, which is directly related to this research. Although Hong et al. [13] tested architectures ranging from ResNet-18 to ViT-B/16, CLIP, GPT-4o-mini, and Gemma-3 on 20,000 images of fast food, the dataset is limited to Western cuisine. When there are scarce or no large annotated food corpora available for a region, transfer learning—taking a pre-trained model from ImageNet and adapting it to the specific domain—is a dominant approach that is used for regional food classification [5, 10]. Gadhiya et al. [4] demonstrated that this is true on a small set of 10 Gujarati food classes, where EfficientNet outperforms in terms of accuracy but uses less data; the fine-grained data assembly and application of pre-trained CNNs are similar to what is done here. Chun et al. [3] also showed that performance is dependent on class distinctiveness of the visual stimuli in a study of 100+ Korean food classes; all three classes of swallow food in Nigeria share texture, color, and form (fufu, semo, and pounded yam). Panduri et al. [10] fused EfficientNet, DenseNet, and MobileNet extractors and compared them to CNN standalone baseline classifiers, using XGBoost and Random Forest classifiers that achieved better performance.

The development of the food classification utility was also found to be applicable to cultural documentation and tourism by Tiwari et al. [5], which reinforces the need for region-specific development. The accuracy was improved when intraclass color variation combined with interclass similarity was a problem, such as with swallow foods, as was proposed by Rollo and Yusiong [9] to use EfficientNetB0 in parallel for processing RGB and HSV color spaces. For the deployment of trained models as mobile applications, the compression, hardware-optimized inference, and user-friendly interfaces for both visually and motor-impaired users are important. Deven and Jacob [6] developed an Android-based calorie estimation application with 81% accuracy on 10 food classes using DenseNet201; this paper extends their work with the ability to output audio for visually impaired users and extends to foods available in Nigeria. For nutritional assessment, Sripada et al. [7] reported an accuracy of 97% using a CNN-SVM hybrid, which surpassed the CNN accuracy rate of 94%, and showed that the approach could be directly applied for visually impaired users who are not able to monitor themselves. This application is embedded in the wider food information landscape powered by AI technologies such as OCR and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

machine learning, as demonstrated by Tejaswi et al. [14], who used OCR in combination with machine learning for food label verification. Reddy et al. [11] have proposed an IoT-based freshness monitoring system with real-time image acquisition, artificial intelligence inference, and mobile application interfaces, with similar structural similarities to the deployment discussed here.

The potential of AI-driven personalized nutrition, demonstrated in biochemical data by Rahman et al. [15], involves users, including the visually impaired, being able to identify foods consumed by themselves, which the study aims to address. Hemalatha and Muthupandi [16] extended the application of the Nigerian swallow food to health and nutrition by converting CNN-classified food objects into nutrition databases for feeding them to the nutrition planner. In their work, Antony and Kumar [8] showed that CNNs are more important than traditional approaches in complex food image feature spaces. In microbial contamination detection, Liu et al. [17] successfully used CNNs to obtain 90% precision and a 91% F1-score, showing their versatility in food safety fields.

The effectiveness of deep learning in food waste detection is demonstrated by Rajagopal et al. [18], who found the deep learning approach significantly superior to conventional algorithms, which supports the use of the CNN architecture. In an effort to put food classification in the context of food safety for visually impaired food consumers, Sateesh et al. [19] formulated a machine learning model for detecting pesticide residues. Rastogi and Ranjana [20] demonstrated that ensemble methods were found to be superior to individual classifiers in the classification of grains and provided suggestions for the improvement of future iterations of the models. Rao et al. [21] performed a multi-class classification on date fruits with a backward feature engineering process that gave interesting concepts in feature selection. For comparison purposes, Reddy and Subbaiah [22] evaluated SVM, KNN, and Random Forest with hyperspectral food quality data. Thomas et al. [23] showed that the effectiveness of EfficientNet-based feature extraction works with limited training data for food commodities, which can be applied directly to the swallow food dataset.

Harishankar et al. [24] developed a framework for classification of Indian food images that combines InceptionNet, MobileNet, VGG16, and ResNet with the explainability methods of SHAP and LIME: With the use of SHAP plots, the visually impaired users can see the visual features that influenced the decision by the AI, while LIME heatmaps can show the region of the image that was predicted by the AI, which is relevant in the context of being unable to verify the decision directly by the visually impaired user.

Tran et al. [25] gave the methodological guidance for data construction of food images and model selection. Sivagami et al. [26] showed this systematic multi-model benchmarking in a similar way as used in the evaluation in this research. In the field of food and health, machine learning has been demonstrated to have wide applicability, with Subedi et al. [27] yielding 90%+ accuracy for food allergy prediction with KNN and ensemble classifiers. Singh et al. [28] have further strengthened the applicability of classification algorithms for dietary disease prevention in vulnerable populations in the context of healthcare. Shevchenko et al. [29] proposed interpretable classification techniques for resource-limited data settings that can be applied to food classification. Milosavljević et al. [30] provided a set of generic pipeline model evaluation frameworks.

CNN transfer learning for food classification has been discussed in the reviewed literature, and it has been found that the mobile deployment is suitable for food identification, dietary monitoring, and nutritional assessment [3, 4, 5]. But there is a critical, persistent need that has not been addressed: a food classification system based on CNN, not to mention a mobile application, has not been created for Nigerian swallow foods. There are no mobile food applications for African food items in the literature, and regional works on cuisines exist for Indian [5], [30], Korean [4], and Gujarati [5]. To the best of the author's knowledge, there is no study that has designed a CNN-based mobile application that has accessibility features in the form of audio for visually challenged users in the Nigerian context. In this work, both gaps are tackled by (i) creating a set of images of swallows from Nigeria; (ii) training a CNN model optimized for this domain; and (iii) using the model to create an audio feedback mobile application for visually impaired people in Nigeria.

METHODOLOGY

This work presents a comparative deep learning pipeline based on six models designed for the classification of foods that can be swallowed by Nigerians, especially for assistive accessibility for visually impaired persons. The pipeline consists of four phases: (i) field data collection and dataset creation, (ii) image pre-processing and augmentation, (iii) training and comparison of the model, and (iv) development of a Flask web application

The six convolutional neural network (CNN) architectures tested are MobileNetV2, LeNet-5, CNN-AFCM (Automatic Food Classification Model), CCFNN (Custom Convolutional Fuzzy Neural Network), SqueezeNet, and ShuffleNetV2. LeNet-5, CNN-AFCM, and CCFNN are trained from scratch, whereas MobileNetV2, SqueezeNet, and ShuffleNetV2 are pre-trained on ImageNet and fine-tuned for the food domain of Nigeria. The winner of the various models proposed is used for on-device deployment, in which it acts as the inference engine of a mobile application that will give real-time audio feedback that will identify the food and its status (fresh or spoiled). The methodological pipeline is presented in Fig. 1 and described below.

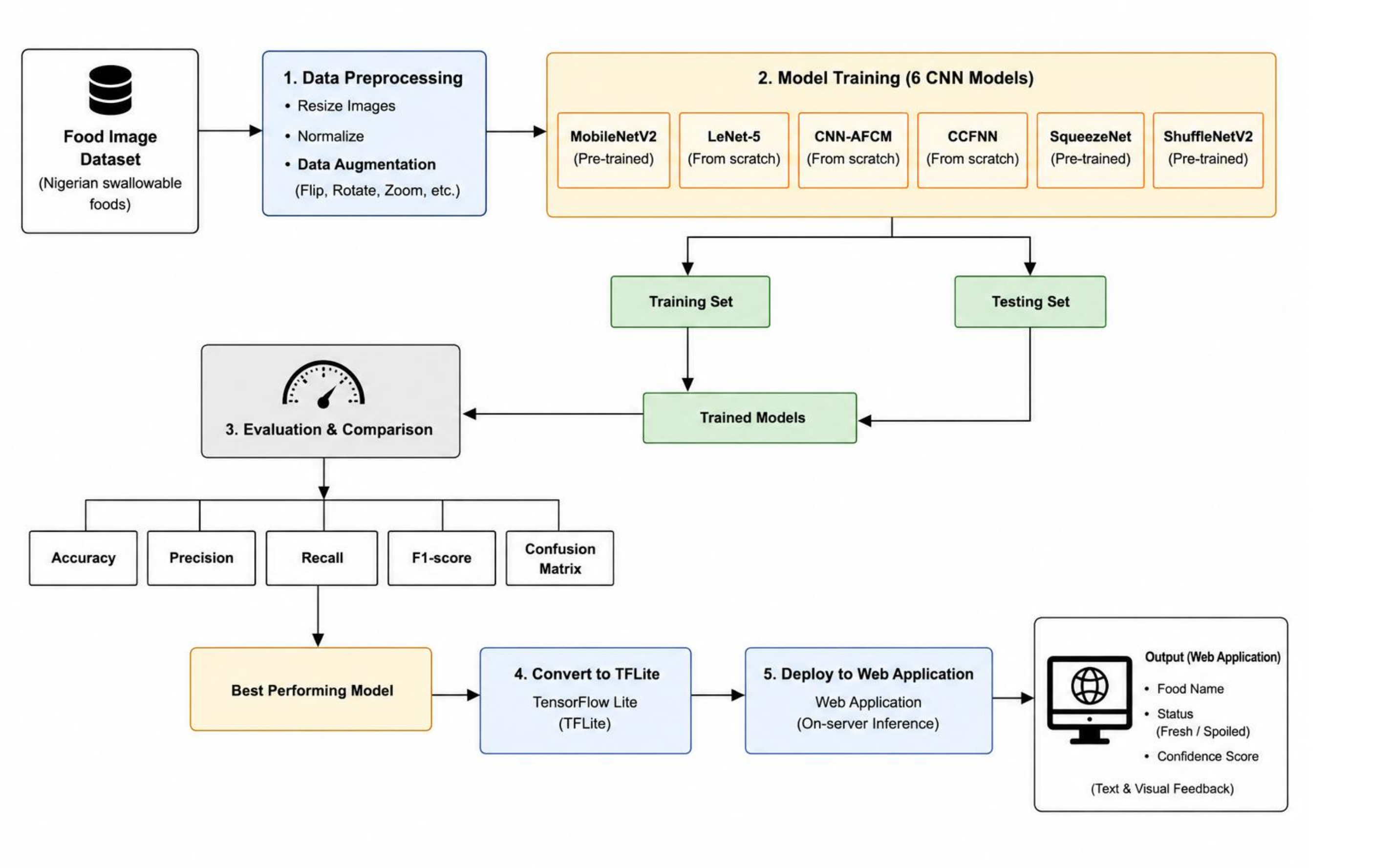

Fig. 1. Architectures of the Proposed System

Dataset Construction

Primary data collection was used to build the data set by the authors. Sixty food-freshness classes (20 × 3) were obtained by taking photographs of 20 food items in Nigeria at three freshness stages: fresh, 4 hours, and 1 day after preparation. This enables the model to identify freshness changes for users with visual impairment who cannot do a visual check. The images were taken in the normal home environment of the Nigerian people using the smartphone camera (Android and iPhone), and there was no studio lighting; this means that the dataset shows the natural variations of background, shadows, and presentation across the Nigerian people. All pictures were captured from a high-down angle that mimics a user's hand holding the phone over a plate. Tables 1 summarize the 60 classes respectively.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Each category includes the following core swallow dishes: fufu, eba, amala dudu, wheat swallow, jollof, white, fried rice, egusi, okro, ewedu, bitter leaf, soko, ugwu, ogi funfun/baba, akara, moimoi, and cooked maize; and two combination dishes of eba/fufu with vegetable and ugwu/ewedu, respectively.

Table 1. Complete List of 60 Food-Freshness Classes

Fresh (Class A)	4-Hour Stage (Class B)	1-Day Stage (Class C)
Fresh Fufu	Fufu (4 Hours)	Fufu (1 Day)
Fresh Eba	Eba (4 Hours)	Eba (1 Day)
Fresh Amala Dudu	Amala Dudu (4 Hours)	Amala Dudu (1 Day)
Fresh Jollof Rice	Jollof Rice (4 Hours)	Jollof Rice (1 Day)
Fresh Fried Rice	Fried Rice (4 Hours)	Fried Rice (1 Day)
Fresh White Rice	White Rice (4 Hours)	White Rice (1 Day)
Fresh Egusi Soup	Egusi Soup (4 Hours)	Egusi Soup (1 Day)
Fresh Okro Soup	Okro Soup (4 Hours)	Okro Soup (1 Day)
Fresh Ewedu	Ewedu (4 Hours)	Ewedu (1 Day)
Fresh Akara	Akara (4 Hours)	Akara (1 Day)
Fresh Beans	Beans (4 Hours)	Beans (1 Day)
Fresh Moimoi	Moimoi (4 Hours)	Moimoi (1 Day)
Fresh Bitter Leaf	Bitter Leaf (4 Hours)	Bitter Leaf (1 Day)
Fresh Ogi Funfun	Ogi Funfun (4 Hours)	Ogi Funfun (1 Day)
Fresh Ogi Baba	Ogi Baba (4 Hours)	Ogi Baba (1 Day)
Fresh Cooked Maize	Cooked Maize (4 Hours)	Cooked Maize (1 Day)
Fresh Ugwu (Fluted Pumpkin)	Efo Ugwu (4 Hours)	Efo Ugwu (1 Day)
Fresh Soko Vegetable	Soko Vegetable (4 Hours)	Soko Vegetable (1 Day)
Fresh Wheat & Soup	Wheat (4 Hours)	Wheat (1 Day)
Fresh Eko & Akara	Eba & Vegetable (Fresh)	Fufu & Vegetable (Fresh)

The dataset comprises 22,823 RGB JPEG images across 60 balanced classes, split 50:30:20 into training, validation, and test sets. The design is balanced with construction, and no class weighting or synthetic oversampling (e.g., SMOTE) was required.

All images were uniformly resized to 224 × 224 × 3, which is the default size for MobileNetV2, SqueezeNet, and ShuffleNetV2 backbones pre-trained on ImageNet. The same settings was used for CCFNN, CNN-AFCM, and LeNet-5. All of the values in the pixels were normalized between 0 and 1. Additionally, for the three

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

pretrained models, ImageNet mean/std normalization (mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]) was applied to the images, in line with their pretraining. All labels were integer encoded (0-59) and then one-hot encoded for training with categorical cross-entropy. To invert and decode the label map during mobile inference, it was stored as a JSON artifact. Augmentation was only used in the training split, on-the-fly, and to replicate real-world hand-held smartphone use. The following six techniques were employed and applied to about 30% of each mini-batch. The splits were not augmented for validation and testing, meaning that evaluation metrics are not based on the performance on augmented data.

CNN Model Architectures

Three architectures were pretrained and fine-tuned on ImageNet and three others were fine-tuned from scratch, the results of which are presented in Table 2.

Table 2. Summary of CNN Architectures Evaluated

Model	Parameters	Trainable Params	Input Size	Architecture Type	Pretrained
MobileNetV2	~2.33 M	~76.9 K (head only, base frozen)	224×224	Depthwise Separable Conv	ImageNet
LeNet-5	~5.62 M	~5.62 M	224×224	Classic CNN (scaled up)	From scratch
CNN-AFCM	~1.28 M	~1.28 M	224×224	Custom multi-block CNN	From scratch
CCFNN	~8.3 K	~8.2 K	224×224	Fuzzy-Gaussian-gated CNN	From scratch
SqueezeNet	~0.77 M	~0.77 M	224×224	Fire modules	From scratch*
ShuffleNetV2	~1.34 M	~1.32 M	224×224	Channel Shuffle Conv	From scratch*

MobileNetV2 [31] employs depth-wise separable convolutions that factor standard convolutions into depth-wise and point-wise convolutions, reducing computation by 8–9× as compared to the equivalent standard convolutions. The number of parameters is low thanks to inverted residual blocks with linear bottlenecks. The backbone of the ImageNet network is preserved, but a head specific to the application is added. Fine tuning is done by progressively unfreezing backbone layers. LeNet-5 [32] is the classical benchmark CNN with a low number of parameters and output layer with 60 classes and trained from scratch using random initialization. CNN-AFCM, introduced in [33] for the classification of Nigerian foods, is made of four convolutional blocks and dense layers with a softmax at the end, with the only difference being the number of classes it outputs, which is 60 instead of 20. CCFNN boosts the convolutional features by introducing fuzzy Gaussian activation, in addition to ReLU, which results in interpretable and human-aligned outputs useful for the trustworthiness of the assistive system. It is currently the lightest model evaluated, which makes it a good model for on-device inference. Trained from scratch. SqueezeNet [34] uses fire modules to reduce to less than 5MB (with almost ~0.77 M parameters) which is good for mobile storage constraints. ShuffleNetV2 [35] proposes to leverage the efficient channel split and channel shuffle operation in order to reuse features without the overhead of group convolutions, resulting in high accuracy per FLOP.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Model Hyperparameter Tuning & Experimental Reproducibility

The model was trained using Kaggle Notebook in Python 3.10 with TensorFlow/Keras up to 50 epochs with early stopping. The hyperparameters were not optimized, but taken from the literature on the task of transfer-learning, including the Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$), the initial learning rate was set to 1e-4, the batch size was set to 64 and the dropout rate was set to 0.3, while the input resolution was set to 224x224x3. The training was done in two stages: First, the pre-trained backbone (MobileNetV2) was warmed up for head-only training (epochs = MAX_EPOCHS/5), then it was tuned at a lower rate of 1e-5. In order to achieve adaptive control, `ReduceLROnPlateau` with the chosen factor of 0.5 and the patience value scaled to MAX_EPOCHS/10 and `EarlyStopping` with patience value scaled to MAX_EPOCHS/5 were used on validation accuracy and the best performing weights were saved by `ModelCheckpoint`. We ensured reproducibility by setting the same kernel augmentation parameters (rotation, zoom, brightness, and shift) for 30% of the batches using Keras layers on the GPU, and by having consistencies in train/val/test folder splits as well as NumPy and TensorFlow random seed values `= 42` everywhere. For fair comparisons of latency, number of parameters, etc., the same mixed precision (float16) training was used on all six models.

The classification accuracy, macro-averaged precision, recall and F1-score, one-vs-rest AUC-ROC, confusion matrix for per-class error analysis, trainable parameter count, and on-device inference latency were used to evaluate models on the held-out test set. To directly measure the effect of augmentation, each of the models was tested in both the with- and without-augmentation settings; the 30% augmentation setting was chosen as a compromise to prevent overfitting.

RESULTS

A total of 22,823 images covering 60 fine-grained classes of foods and soups in the Nigerian context were captured under two storage-time conditions (e.g., 1-day and 4-hour variants) to simulate realistic food freshness scenarios. The data was split into training, validation, and test data subsets. Most classes were fairly well represented, but some of the classes in the "Fresh" category (e.g., Fresh Ebia and Vegetable, Fresh Wheat and Soup) were underrepresented to some extent—as few as 20-30 test images—which is a point that was brought up again in the discussion on class-level confusion below.

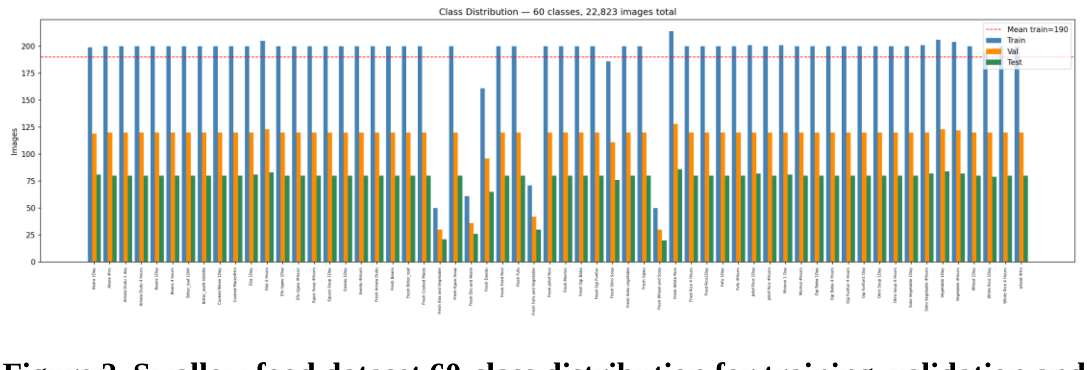

Figure 2. Swallow food dataset 60-class distribution for training, validation and test sets: (n = 22,823).

The various CNN architectures were trained and tested on the held-out test set for accuracy, macro-averaged precision, recall, F1-score and one-vs-rest AUC-ROC. All six models are compared in terms of their performance and are summarised in Table 3.

Table 3. Six CNN models are evaluated in the 60-class food test set for swallow.

Model	**Accuracy**	**Precision (macro)**	**Recall (macro)**	**F1-score (macro)**	**AUC-ROC**	**Params (M)**
MobileNetV2	**0.9978**	**0.9979**	**0.9979**	**0.9979**	**1.0000**	**2.33**

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

LeNet-5	0.9943	0.9946	0.9937	0.9940	1.0000	5.62
ShuffleNetV2	0.9895	0.9902	0.9893	0.9896	1.0000	1.34
SqueezeNet	0.9699	0.9697	0.9693	0.9690	0.9994	0.77
CNN-AFCM	0.8760	0.9008	0.8799	0.8615	0.9993	1.28
CCFNN	0.2354	0.1838	0.2216	0.1711	0.9325	0.01

Despite the moderate number of parameters compared with the other architectures being assessed (2.33M), MobileNetV2 had the highest value of test accuracy (99.78%), a macro F1-score of 0.9979, and an AUC-ROC of 1.0000. LeNet-5 with 99.43% accuracy and ShuffleNetV2 with 98.95% accuracy were very close behind and both performed perfectly on the AUC-ROC score. SqueezeNet achieved an acceptable accuracy of 96.99% and the lowest parameter requirement of 0.77M out of the conventionally trained models. CNN-AFCM was the trailing group with 87.60% accuracy, while CCFNN obtained much inferior performance for all measurements (23.54% accuracy; F1-macro = 0.1711).

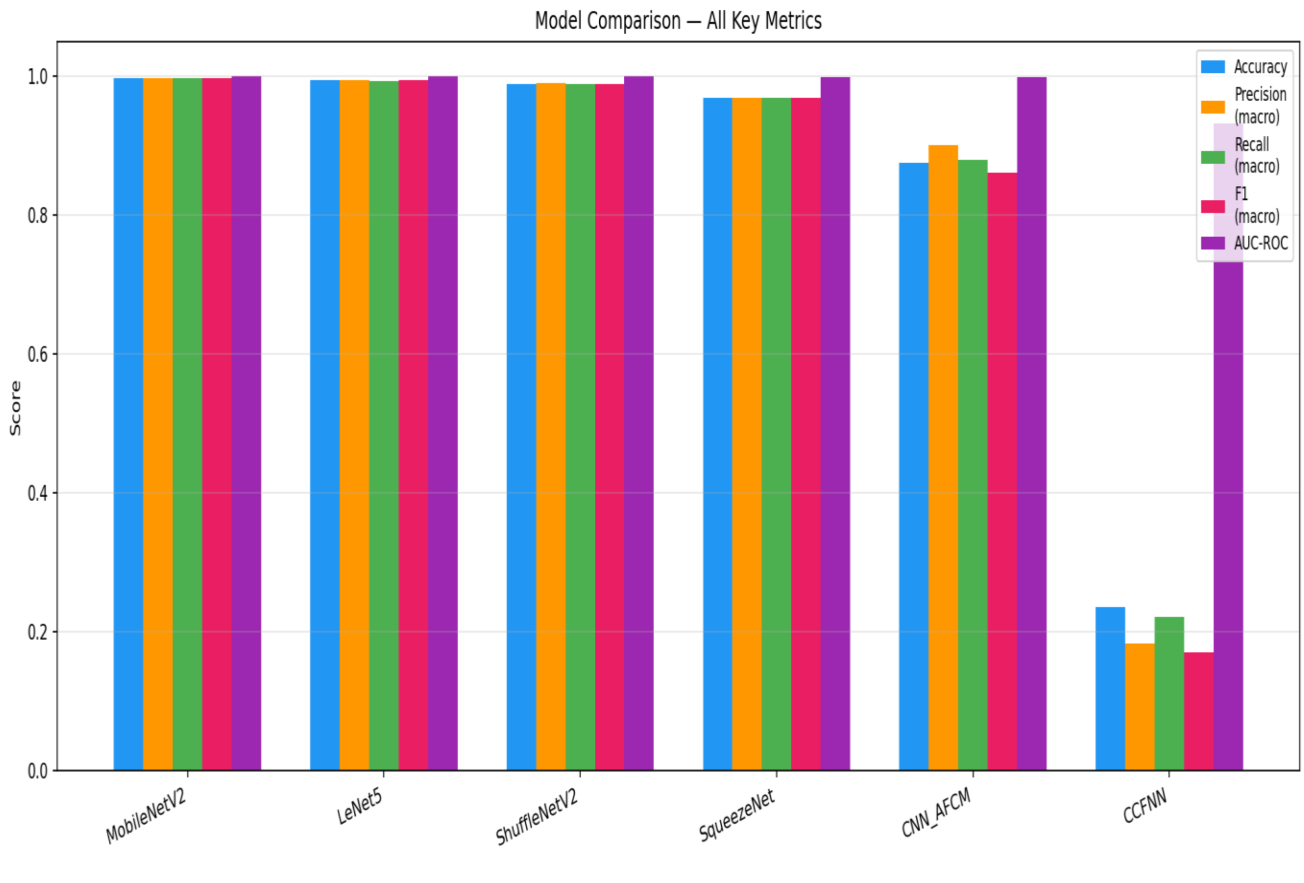

Figure 3. Side-by-side comparison of accuracy, macro precision, macro recall, macro F1-score, and AUC-ROC across all six architectures.

Best-Performing Model: MobileNetV2

For MobileNetV2 (Figure 4), the confusion matrix displays a near-perfect diagonal line across each of the 60 classes that represent the accuracy of the classification, with little to no intra-class confusion. The 17 misclassifications that were found were grouped within the same pairs of “Fresh” categories, as expected due to the relatively lower number of samples identified for these classes (Section 4.1), which are visually and semantically similar (e.g., Fresh Ebia and Vegetable versus Fresh Egusi Soup, and Fresh Eko and Akara versus Fresh Fufu and Vegetable).

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

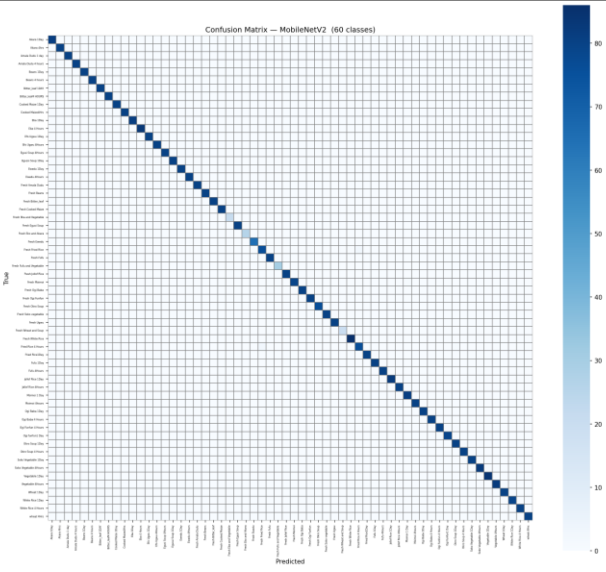

Figure 4. Confusion matrix for MobileNetV2 across all 60 classes.

MobileNetV2 (Figure 5) demonstrates a rapid initial convergence followed by a transient increase in the validation loss and corresponding drop in the validation accuracy at epochs 10-11, where a fine-tuning event occurred (in this case, layers in the backbone were unfrozen and/or the learning rate was scaled up). Afterward, the model gradually improved, with training and validation losses stabilizing to low, consistent levels by epoch 50, and no signs of overfitting.

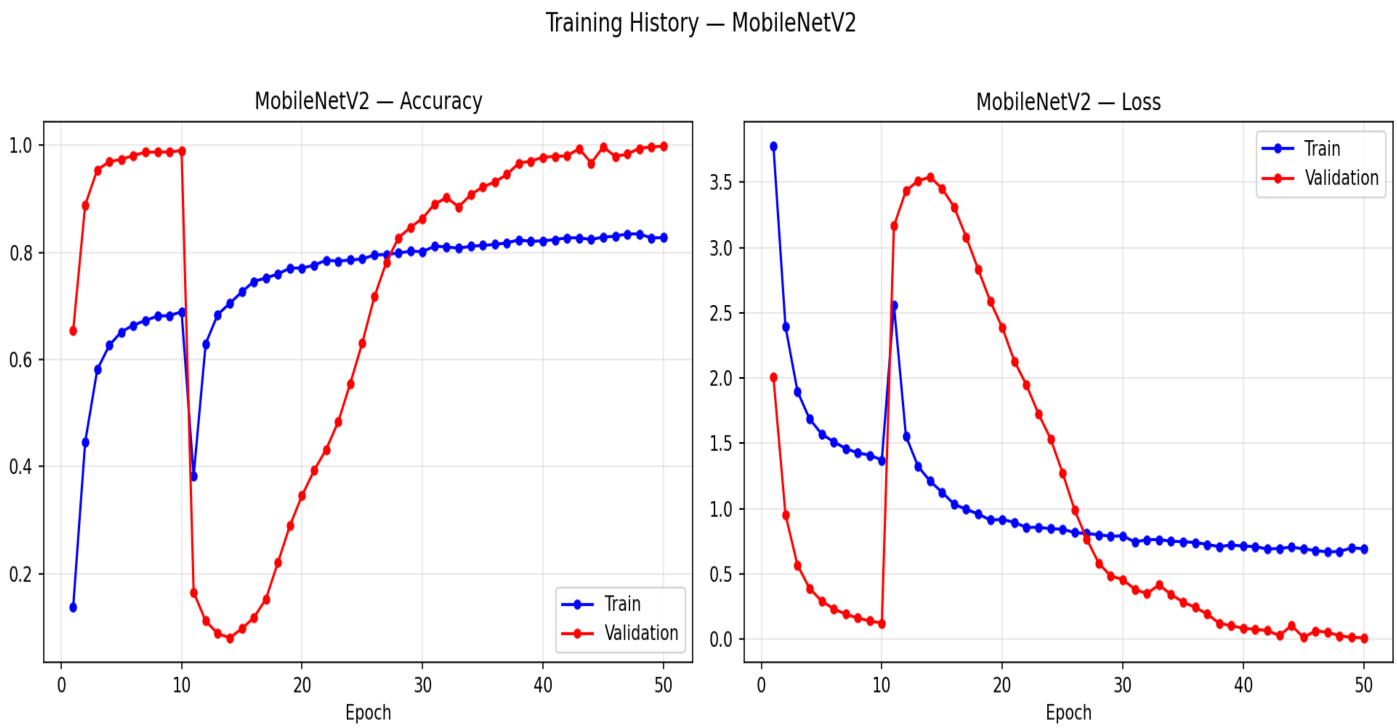

Figure 5. MobileNetV2 training and validation accuracy and loss curve for 50 epochs

Underperforming Model: CCFNN

Compared with the other five architectures, CCFNN could not learn the classification task well: It obtained a test accuracy of 23.54% and a macro F1-scores of 0.1711. The confusion matrix (Figure 6) demonstrates that predictions are widely distributed away from the diagonal, with the model converging towards a restricted number of classes, rather than the entire 60-class space.

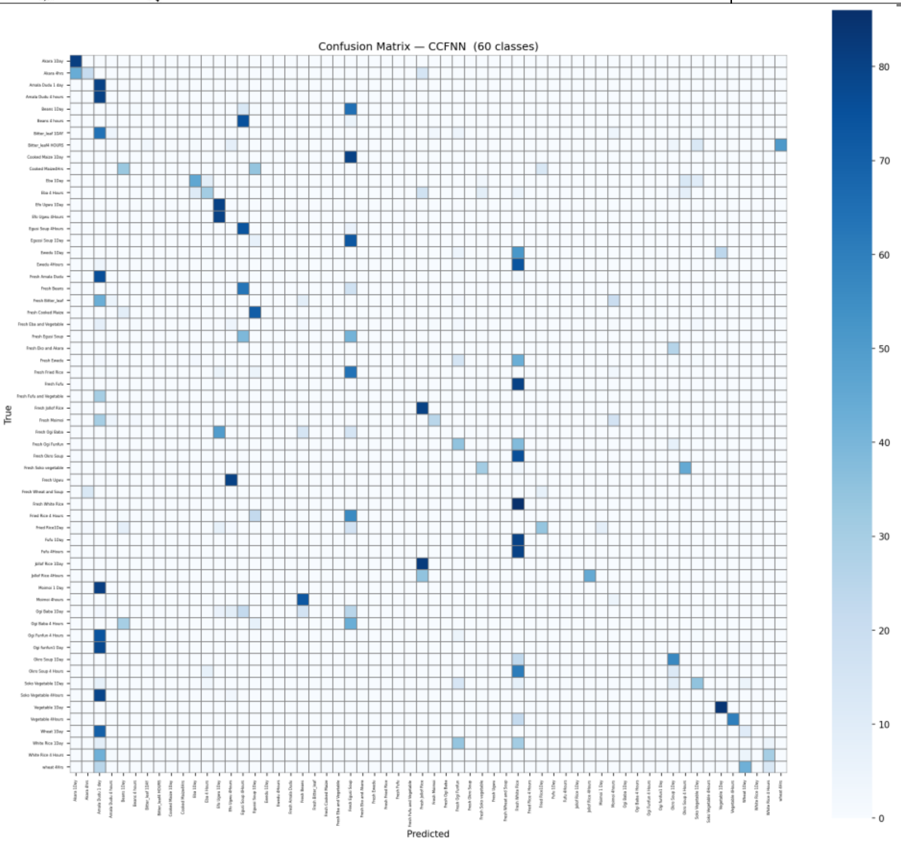

Figure 6. Confusion matrix for CCFNN, showing widespread off-diagonal misclassification.

This is substantiated by the training history (Figure 7: training and validation loss do not converge towards a low value after 50 epochs, and both the training and validation loss stay at a relatively high value (~3.1–3.5), and training accuracy increases slowly to ~13% whilst validation accuracy ranges between ~19% and ~23%. While this is a data-handling or evaluation artefact, this pattern, also observed for other architectures evaluated (Figure 8), seems to indicate that the CCFNN has not enough representational capacity to model the distinctions it needs to make for fine-grained visual classification of the 60 food classes, having much less than 10,000 trainable parameters. Interestingly, CCFNN's AUC-ROC (0.9325) is significantly higher than the accuracy or F1-score, which suggests that although the model still has some capacity to rank correct classes above incorrect ones probabilistically, this is not being reflected in predictable and correct top-1 predictions. This difference between ranked and decision-based metrics adds to the argument that this is not necessarily a flawed learning process, but rather an underfitting one.

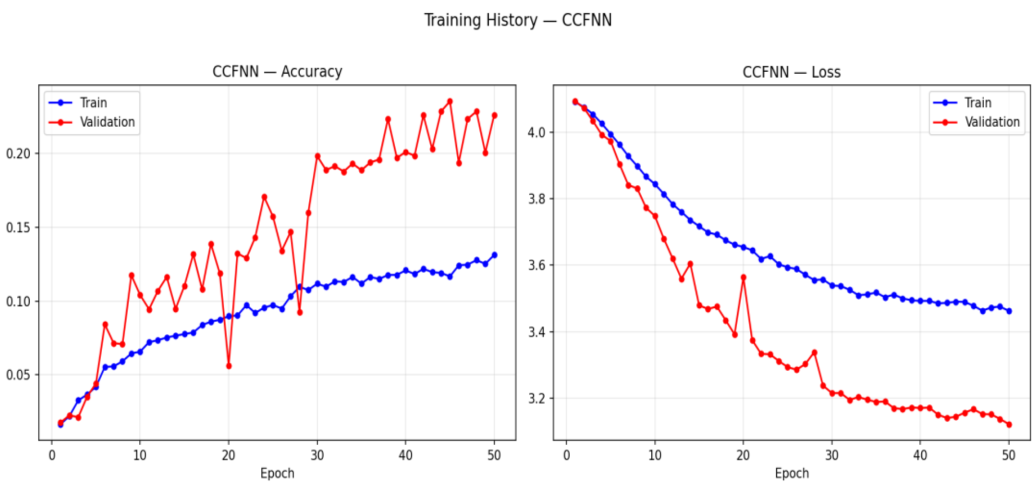

Figure 7. Training and validation accuracy and loss curves of CCFNN, with a plateau indicative of underfitting.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

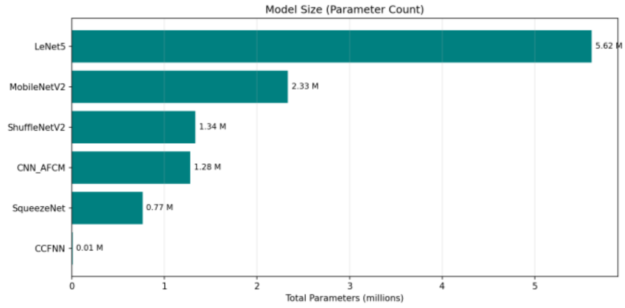

Figure 8. Number of trainable parameters for each architecture (in millions).

The test accuracy versus the mean inference latency per image is shown in figure 9 for the five conventionally performing models and CCFNN. Among the models compared, MobileNetV2 is the most favourable in terms of accuracy, but also has the highest latency (86 ms/image). SqueezeNet and CNN-AFCM have moderate accuracy, which, however, provides faster inference time, about 80 ms/image, while CCFNN has the lowest latency; however, it is not suitable for deployment due to its low accuracy. The results indicate that either ShuffleNetV2 or SqueezeNet might be a better option for resource-constrained applications (e.g., a food-recognition app on a mobile phone or edge device) than MobileNetV2, and, in cases where the primary goal is to achieve maximum classification accuracy, MobileNetV2 is the best option.

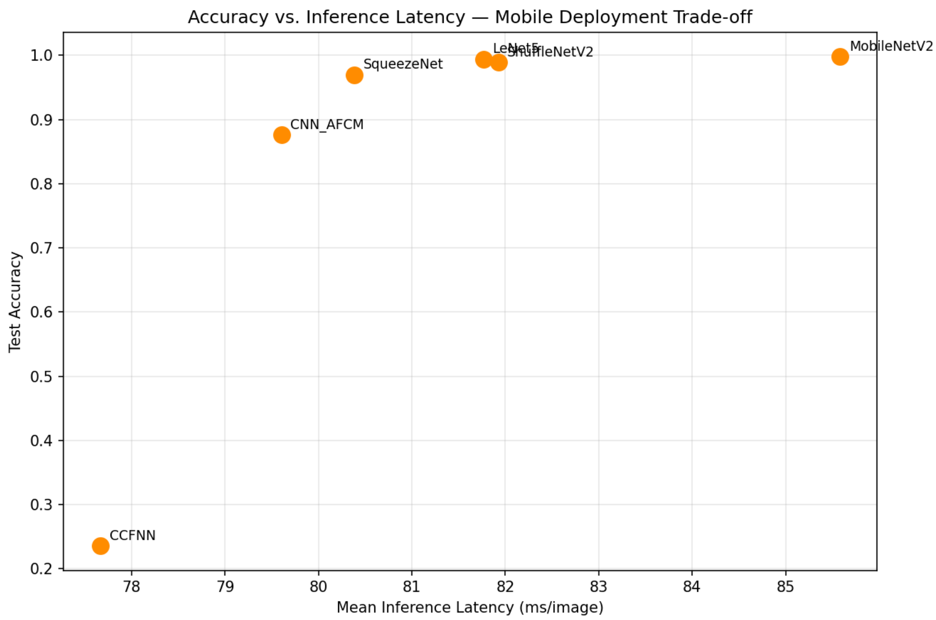

Figure 9. Test accuracy versus mean inference latency (ms/image), illustrating the accuracy–efficiency trade-off across architectures.

DISCUSSION

To the best of our knowledge, the comparative evaluation presented here shows that lightweight, mobile-oriented CNN architectures, especially MobileNetV2, are well suited for fine-grained food classification of Nigerian swallowed food, achieving near-perfect classification accuracy for 60 classes despite the visual similarity between some foods and their “fresh” counterparts. The AUC-ROC scores are very high (above 0.9993) for five of the six architectures, suggesting that the dataset, as designed, is well-separable over the space of its features for models of sufficient representational power. The limited parameter capacity of CCFNN (around 0.01M trainable parameters) is not a reason for the poor performance compared to the other five architectures, but rather the dataset quality and labeling, as can be seen in the training curves (i), where the CCFNN shows a convergence pattern that is not very consistent and not very complete; and in the confusion matrix (ii), where the misclassifications are distributed and not systematic; and in the difference between its AUC-ROC and its accuracy/F1 scores (iii), which are marked. This result is in line with previous work reporting reduced

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

performance on fine-grained high-class-count tasks for extremely light architectures and highlights the fact that the reduction in the number of parameters to handle deployment efficiency is a meaningful trade-off for accuracy if the parameter count is reduced too much. The results of CCFNN are reported instead of being excluded due to their meaningful representation of the lower bound of the capacity–accuracy trade-off investigated in this study. Overall, these findings suggest that, for the food classification task in the Nigerian swallow application proposed in this paper, the MobileNetV2 architecture is the recommended choice between it and ShuffleNetV2 and SqueezeNet in terms of accuracy and the number of parameters, respectively, which are of moderate size suitable for mobile deployment.

The limitation of the study is that usability testing with visually impaired users was not conducted in this phase and is deferred to future work.

Application Interface

The model trained using the MobileNetV2 architecture was then deployed in a web application (Naija Food Eye) using Flask as a web application framework, which was created as an assistive technology for the visually impaired user. It was initially planned for on-device Flutter deployment; due to some reasons, the final system was deployed as a Flask web application. The homepage (Figure 10) shows headline model statistics (accuracy 99.78%, AUC-ROC 1.0000, 60 classes) and upload of images panel. Once a user takes or uploads a picture of food (Figure 11), the system then responds with the top predicted class, a confidence score, the top-5 alternative predictions, a freshness label (Fresh, 4 Hours Old, 1 Day Old), and inference latency (Figure 12). In the example displayed, the model could accurately guess "Amala Dudu 1 day" with a 99.9% confidence in just about 12.5 seconds.

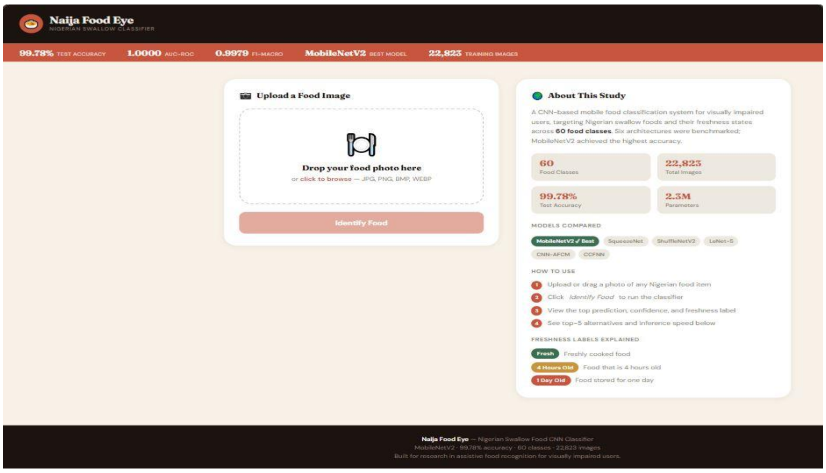

Figure 10. Naija Food Eye home page with model statistics and image upload page.

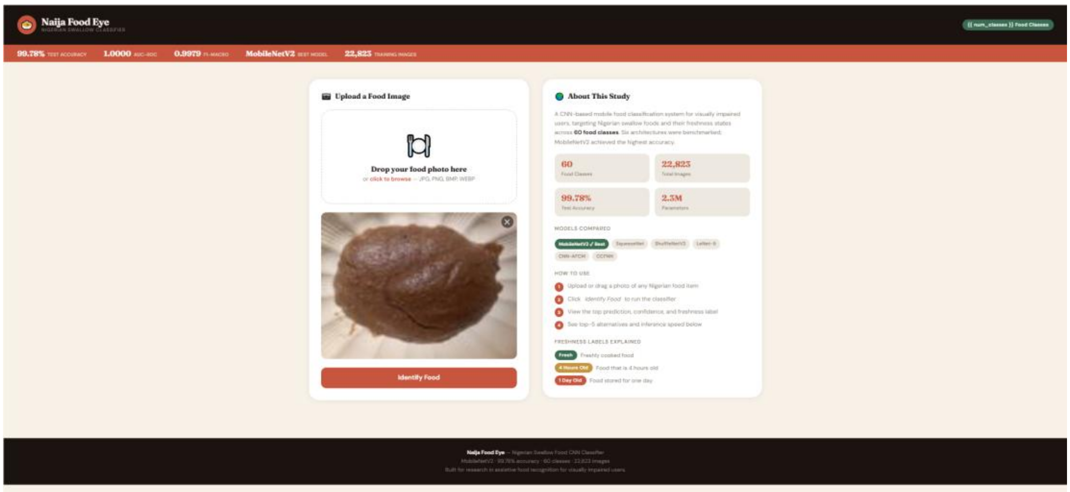

Figure 11. Food image upload step prior to classification.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

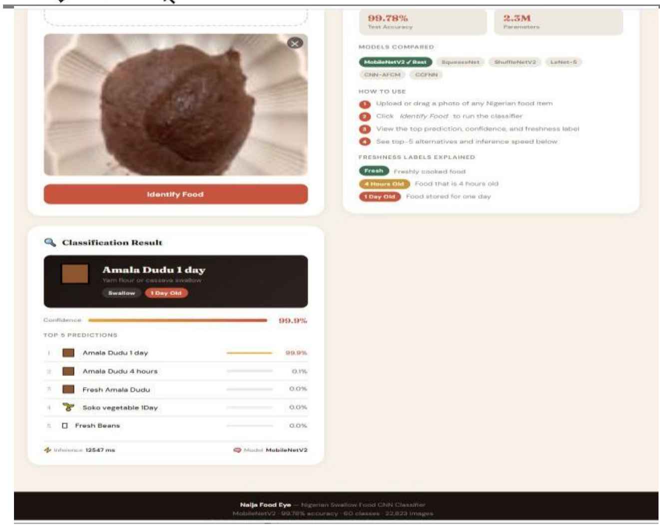

Figure 12. The result panel for classification with top prediction, confidence, top-5 alternatives, and inference time

CONCLUSION

This study compared the performance of six convolutional neural network architectures; MobileNetV2, LeNet-5, ShuffleNetV2, SqueezeNet, CNN-AFCM, and CCFN on the fine-grained classification of 60 swallow food classes and 22,823 swallow food images comprising the freshness states (fresh, 4 hours old, 1 day old) and food identity. MobileNetV2 outperformed other architectures, with a test accuracy of 99.78%, a macro F1-score of 0.9979, and an AUC-ROC of 1.0000, with a moderate number of parameters. LeNet-5 and ShuffleNetV2 succeeded closely with 98.9% accuracy, and SqueezeNet had a good efficiency trade-off due to its smaller model size. The performance of CNN-AFCM was moderate, while that of CCFNN underperformed because of its lack of representational capacity, which was explained by convergence and confusion-matrix analysis and not considered an anomaly.

This best-performing model, namely, MobileNetV2, was successfully deployed in a web application using Flask, where the visually impaired users are able to find out what Nigerian swallow foods are being taken based on one image provided by them and also able to determine the freshness of the foods based on the same image provided to them. This proved the feasibility of the proposed system as an assistive tool for identification and freshness assessment of Nigerian swallow foods using a single image provided by the visually impaired user. All the results together show that lightweight, mobile CNN architectures can enable high-accuracy, fine-grained food recognition in a culturally specific domain with limited previous research focus and demonstrate the accuracy–capacity compromises needed when choosing an architecture for real-world use.

Future Work

There are Several Directions That can Further Extend This Work. Expansion of the Dataset to Cover a Broader Selection of Dishes in Nigeria Other than the Swallow Foods and Their Soups, as Well as More Freshness Intervals than the 3 Modelled in This Work, to Increase Generalisability of the Work. Second, On-Device Deployment and Benchmarking on Real Mobile or Embedded Hardware (Not Through a Latency Estimate in a Development Environment) Would Best Characterize Real-World Inference Performance for the Target at Use Case. Third, Should Ultra-Lightweight Models be Able to Match the Performance of the Larger Models Within CCFNN, Future Research Will Investigate the Potential for Boosting the Capacity of the Parameters or the Architecture (Such as Adding More Convolutional Blocks or Attention Mechanisms) Without Sacrificing the Efficiency Benefit. Fourthly, Using Audio Feedback and Voice-Guided Interaction with the Application Naija Food Eye Will Enhance the Accessibility of the Application for Visually Impaired Individuals, Not Only in the Visual Section of the Result Panel. Last but Not Least, Although Not a Topic of the Present Invention, It is Appreciated That a Field Test Would be Performed with Visually Impaired Subjects in Actual Dining and Food

Storage Environments to Gain Useful Usability-Based Feedback to Complement the Technical Performance Results Reported Herein.

ACKNOWLEDGEMENTS

The authors would like to thank the Federal University of Technology, Ilaro (FUTI), Ogun State, Nigeria, for the support of this study under the Year 2024–2025 TETFund Intervention in Research Project (RP), Batch 10 with reference number AD/RDC/38/VOL.II/132, Tertiary Education Trust Fund (TETFund) Institution Based Research (IBR) Intervention. The authors are also grateful of the financing support of TETFund and the Federal Government of Nigeria for this study.

Funding Information

The research work has been funded by the Tertiary Education Trust Fund (TETFund) Institution-Based Research (IBR) Intervention, Research Project (RP) Batch 10, Year 2024 – 2025, Federal University of Technology, Ilaro (FUTI), Ogun State, Nigeria, under grant reference number: AD/RDC/38/VOL.II/132.

REFERENCES

1. World Health Organization (WHO), "World Report on Vision," WHO Press, Geneva, Switzerland, 2019. [Online]. Available: https://www.who.int/publications/i/item/9789241516570
2. S. Memiş, B. Arslan, O. Z. Batur, and E. B. Sönmez, "A Comparative Study of Deep Learning Methods on Food Classification Problem," in Proc. 2020 IEEE Conf. Signal Processing and Communication Applications, Istanbul, Turkey, 2020, pp. 1–4. doi: 10.1109/SIU49456.2020.9302375
3. M. Chun, H. Jeong, H. Lee, T. Yoo, and H. Jung, "Development of Korean Food Image Classification Model Using Public Food Image Dataset and Deep Learning Methods," IEEE Access, vol. 10, pp. 128732–128743, Dec. 2022. doi: 10.1109/ACCESS.2022.3227796
4. J. Gadhiya, A. Khatik, S. Kodinariya, and D. Ramoliya, "Classification of Regional Food Using Pre-Trained Transfer Learning Models," in Proc. 7th IEEE Int. Conf. Electronics, Communication and Aerospace Technology (ICECA), 2023, pp. 1–6. doi: 10.1109/ICECA58529.2023.10395249
5. R. G. Tiwari, V. Gautam, A. K. Jain, and V. Sharma, "A Study on Food Classification and its Significance in Promoting Tourism and Boosting the Economy," in Proc. 2nd IEEE Int. Conf. Augmented Intelligence and Sustainable Systems (ICAISS), 2023, pp. 1–6. doi: 10.1109/ICAISS58487.2023.10250473
6. S. S. Deven and P. M. Jacob, "An Android App Based Food Calorie Estimator Using Machine Vision," in Proc. 3rd IEEE Int. Conf. Advances in Computation, Communication and Information Technology (ICAICCIT), 2025, pp. 1–5. doi: 10.1109/ICAICCIT68829.2025.11434291
7. N. K. Sripada, S. C. Challa, and S. Kothakonda, "AI-Driven Nutritional Assessment Improving Diets with Machine Learning and Deep Learning for Food Image Classification," in Proc. IEEE Int. Conf. Self Sustainable Artificial Intelligence Systems (ICSSAS), 2023, pp. 1–6. doi: 10.1109/ICSSAS57918.2023.10331787
8. A. M. Antony and R. S. Kumar, "A Comparative Study on Predicting Food Quality using Machine Learning Techniques," in Proc. 7th IEEE Int. Conf. Advanced Computing and Communication Systems (ICACCS), 2021, pp. 1–5. doi: 10.1109/ICACCS51430.2021.9441743
9. S. Jr. D. Rollo and J. P. T. Yusiong, "A Two-Color Space Input Parallel CNN Model for Food Image Classification," in Proc. IEEE Int. Conf. Information Technology Research and Innovation (ICITRI), 2024, pp. 138–143. doi: 10.1109/ICITRI62858.2024.10698941
10. B. Panduri, N. Vullam, R. S. S. R. Battula, V. D. Babu, V. S. Desanamukula, and A. Lakshmanarao, "Food Classification Using a Hybrid Framework with Transfer Learning and Machine Learning Models," in Proc. IEEE Int. Conf. Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), 2025, pp. 1–6. doi: 10.1109/AMATHE65477.2025.11081177

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

11. E. K. Reddy, T. S. P. Balaji, and K. K. Manivannan, "AI Based Food Freshness Tracking System with Shelf-Life Prediction," in Proc. 5th IEEE Int. Conf. Intelligent Technologies (CONIT), 2025, pp. 1–5. doi: 10.1109/CONIT65521.2025.11166784
12. J. Sultana, B. M. Ahmed, M. M. Masud, A. K. O. Huq, M. E. Ali, and M. Naznin, "A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques," IEEE Access, vol. 11, pp. 45904–45930, May 2023. doi: 10.1109/ACCESS.2023.3274475
13. A. Hong, J.-S. Ma, and S. Bellur, "Comparative Analysis of AI Models for Fast Food Image Classification," in Proc. IEEE Int. Conf. Electrical and Computer Engineering Researches (ICECER), 2025, pp. 1–6. doi: 10.1109/ICECER65523.2025.11401344
14. K. Tejaswi et al., "Automating Nutritional Claim Verification: The Role of OCR and Machine Learning in Enhancing Food Label Transparency," in Proc. IEEE Int. Conf. IoT Based Control Networks and Intelligent Systems (ICICNIS), 2024, pp. 1–5. doi: 10.1109/ICICNIS64247.2024.10823177
15. R. Rahman, M. Hassan, A. F. Tasnim, and J. Ferdousmou, "AI-Powered Personalized Nutrition Based on Biochemical Data: An Exploration of Machine Learning Applications," in Proc. IEEE Int. Conf. Contemporary Computing and Communications (InC4), 2025, pp. 1–6. doi: 10.1109/InC465408.2025.11256407
16. M. Hemalatha and G. Muthupandi, "A Comprehensive Framework for Nutrition Analysis and Ingredient Substitution using Machine Learning," in Proc. IEEE Int. Conf. Data Science, Agents and Artificial Intelligence (ICDSAAI), 2025, pp. 1–6. doi: 10.1109/ICDSAAI65575.2025.11011623
17. X. Liu et al., "Automatic Detection Model of Food Microbial Contamination Combined with Image Processing and Pattern Recognition," in Proc. IEEE Int. Conf. Telecommunications and Power Electronics (TELEPE), 2024, pp. 1–5. doi: 10.1109/TELEPE64216.2024.00121
18. R. Rajagopal and K. Chaithanya, "Comparative Study of Machine Learning and Deep Learning Models for Food Waste Detection Using Image Processing," in Proc. IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS), 2025, pp. 1–5. doi: 10.1109/SCEECS64059.2025.10941059
19. N. Sateesh, N. Rawat, W. Patel, and H. Patil, "A Classification Model for Pesticide Residues in Food Based on Machine Learning Methods," in Proc. 5th IEEE Int. Conf. Innovative Trends in Information Technology (ICITIIT), 2024, pp. 1–5. doi: 10.1109/ICITIIT61487.2024.10580640
20. S. Rastogi and Ranjana, "Comparative Analysis of Machine Learning Algorithms for Wheat Grain Classification using Ensemble Learning Approaches," in Proc. 14th IEEE Int. Conf. System Modeling and Advancement in Research Trends (SMART), 2025, pp. 1–6. doi: 10.1109/SMART66937.2025.11389335
21. B. S. N. Rao et al., "An Innovative Machine Learning System for Optimal Multi-Class Classification of Date Fruits by Leveraging Backward Feature Engineering to Promote Sustainable Agriculture," in Proc. IEEE Int. Conf. Integrated Intelligence and Communication Systems (ICIICS), 2024, pp. 1–6. doi: 10.1109/ICIICS63763.2024.10859468
22. S. Reddy and V. Subbaiah, "Comparative Analysis of Hyperspectral Imaging with Machine Learning Assessment of Crop Health and Honey Quality," in Proc. 2nd IEEE Int. Conf. Recent Advances in Information Technology for Sustainable Development (ICRAIS), 2024, pp. 1–5. doi: 10.1109/ICRAIS62903.2024.10811699
23. E. S. Thomas et al., "EfficientNet-Based Feature Extraction and Ensemble Learning for Tomato Quality Classification," in Proc. 3rd IEEE Int. Conf. Intelligent Systems, Advanced Computing and Communication (ISACC), 2025, pp. 1–6. doi: 10.1109/ISACC65211.2025.10969232
24. C. Harishankar, S. Anoop, and K. S. Niranjana, "An Explainable Hybrid Learning Model for Indian Food Image Classification," in Proc. 15th IEEE Int. Conf. Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1–6. doi: 10.1109/ICCCNT61001.2024.10724730
25. Q.-L. Tran, G.-H. Lam, Q.-N. Le, T.-H. Tran, and T.-H. Do, "A Comparison of Several Approaches for Image Recognition used in Food Recommendation System," in Proc. IEEE Int. Conf. Communication, Networks and Satellite (COMNETSAT), 2021, pp. 1–6. doi: 10.1109/COMNETSAT53002.2021.9530793
26. S. Sivagami et al., "Automated Digitally Modulated Signal Recognition and Classification using Machine Learning with Multimodal Information," in Proc. 2nd IEEE Int. Conf. Augmented

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Intelligence and Sustainable Systems (ICAISS), 2023, pp. 1–5. doi: 10.1109/ICAISS58487.2023.10250687

27. S. Subedi et al., "Enhancing Allergy Prediction Accuracy Through Machine Learning and ProteinBERT," in Proc. IEEE Int. Conf. Machine Learning and Applications (ICMLA), 2024, pp. 1–6. doi: 10.1109/ICMLA61862.2024.00068
28. A. Singh et al., "Diabetes Prediction System Using Machine Learning," in Proc. IEEE Int. Conf. Advances in Computation, Communication and Information Technology (ICAICCIT), 2023, pp. 1–5. doi: 10.1109/ICAICCIT60255.2023.10466034
29. V. Shevchenko, A. Lukashevich, and Y. Maximov, "Climate Change Impact on Agricultural Land Suitability: An Interpretable Machine Learning-Based Eurasia Case Study," IEEE Access, vol. 12, pp. 14000–14020, Jan. 2024. doi: 10.1109/ACCESS.2024.3358865
30. S. Milosavljević et al., "Exploiting Machine Learning for Industrial Defrost Cycle Characterization," in Proc. 33rd IEEE Telecommunications Forum (TELFOR), 2025, pp. 1–5. doi: 10.1109/TELFOR67910.2025.11314454
31. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
32. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
33. J. R. Olasina and O. H. Aliu, "Development of convolutional neural network (CNN)-based automatic food classification (AFC) model for visual impairments of Nigerians," in Applied Mathematics, Modeling and Computer Simulation. Amsterdam, The Netherlands: IOS Press, 2024, pp. 2–10, doi: 10.3233/ATDE240774.
34. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5MB model size," arXiv preprint arXiv:1602.07360, 2016.
35. N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in Proc. Eur. Conf. Computer Vision (ECCV), Munich, Germany, 2018, pp. 116–131, doi: 10.1007/978-3-030-01264-9_8.