International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 14, Issue 11 [November 2025] PP: 95-101

Development of Machine Learning Model for Classification of Annual Weeds in South Western Nigeria.

Aworetan F.A and Oluwadare S. A
Department of Computer Science Federal University of Technology Akure

Abstract
The weeds in South Western Nigeria are another significant constraint to crop production resulting in loss of a large yield and economic impacts. The traditional weed identification and classification has been inaccurate in general and time consuming and labor intensive. Machine learning opens a prospect of improving the system of automation of the weed classification process and helps to introduce precision agriculture, along with sustainable cultivation of crops. The journal article offers the construction of a machine learning model to categorize weeds in South Western Nigeria, where a collection of images of the common weed species in the region was used. The model was exceptionally precise in identifying the character of the weeds, thus signifying that machine learning can be used to advance the control of the weeds in the region.
Keywords: *weed classification, machine learning, image recognition, south western, convulational neural network,*

Date of Submission: 04-11-2025 Date of acceptance: 14-11-2025

I. Introduction

The problem of weeds is a huge issue facing crop production in South Western Nigeria where agricultural activities play a major role in the economy. Conventional approaches to identifying and characterising weeds are based on both manual observation and experience and these approaches can be time-consuming, labour intensive and are also subject to error. Machine learning offers a promising solution for automating weed classification, enabling farmers to make informed decisions about weed management and reducing the environmental impact of herbicide use.

II. Literature Review

This journal provides a background of existing K-means clustering solution solving problems in computer, Annual Weed Identification Problem (AWIP), Annual Weed Classification Problem (AWCP), Unknown Weed Identification and Classification Problem (UWICP) and K-means Clustering Algorithm. It gives a review of related research works in literature on solved problems related to weed identification and classification. Also, it provides a summary description of methodologies and solutions which may be extraction of clustering sample data set, generic variables to all annual weed plants. The growth of the annual weeds which compete with crops in terms of water, nutrients, sunlight, and space critically affects the agricultural productivity of the developing countries especially in sub-Saharan Africa. South west Nigeria is a region where maize, cassava and yams are cultivated and the weeds are growing uncontrollably during rainy season resulting in huge loss of yield. Conventionally, weed management was largely dependent on manual labor and herbicides, neither of which is economically and ecologically viable.

New innovation in data mining, machine learning, and computer vision have offered new solutions to automation of the process of detecting and classifying weeds. The strategies can enhance the accuracy and effectiveness of weed control methods. An example of this is the CWD30 dataset offered by Ilyas et al. (2023) that has more than 219,000 annotated images of weed and crops to train deep learning models in various field scenarios. Equally, Islam et al. (2025) showed that the multi-stage detection of RetinaNet is superior to the DETR when monitoring the growth of the weed with a classification accuracy of more than 90 per cent.

Despite the fact that some of the models have been very high-performing under controlled conditions, their acclimatization to the local conditions is yet to be achieved. Research like that by Damilare et al. (2023) and Owoweye et al. (2024) have made crucial steps in implementing the YOLO-based deep learning model to categorize the species of maize and weed in agricultural fields in Nigeria. Nonetheless, there are constraints that include local data constraints, the differentiation of the weed species in different ecological areas and the affordability of solutions to such challenges that smallholder farm owners can afford.

The agricultural sector is still a vital industry to Nigeria economy with more than 60 percent of the population engaged in it and enhancing food security and rural development. Among the myriad of issues that plague agricultural production, one can single out the issue of weeds as one of the main causes of low crop yields in particular agricultural production that occur in rain-fed agricultural production systems that are common in Southwest Nigeria. The Tridax procumbens, Ageratum conyzoides, Chromolaena odorata and Euphorbia heterophylla are examples of some of the every-year weeds that compete with the staple crops, like maize, cassava and yam among other crops, in these areas; that is, they struggle over essential resources such as sunlight and nutrient and water. Hand weeding of these weeds is both tedious, time-consuming, and frequently inefficient resulting in low production in the farm and high production cost (Adeleye et al., 2022).

In the recent past, data mining and machine learning (ML) have offered new methods of automating the process of identifying and classifying weeds. This type of smart technology has the potential to analyze massive amounts of agricultural intelligence and sensor data to differentiate between crops and weeds species, enabling agricultural machines to weed more precisely and apply herbicides to specific areas instead of being applied broadly. To illustrate, Ilyas et al (2023) developed a massive dataset (CWD30) with more than 219,000 labeled photographs of weeds and crops to assist in deep learning classification in various farm conditions. This project demonstrates the potential that machine learning has to continue to develop to allow the automation of agricultural diagnostics and yield optimization.

Studies conducted in the context of the world market have demonstrated that object detection and classification systems like YOLOv5, RetinaNet, and ResNet can reach a classification accuracy of more than 90% in detecting various types of weeds (Islam et al., 2025; Silva et al., 2024). Nevertheless, such models usually use high-quality and annotated datasets which are not common in less developed countries such as Nigeria. Moreover, WeedNet, which is a global to local AI model, was proposed by Shen et al. (2025) to enable the pre-trained classifiers to be localized to implement them in real-time. They found that in regional applications, local fine-tuning of models yielded a large classification accuracies, up to the point of more than 97.4 percent accuracy. These results prove the idea of creating localized datasets and models that will consider the ecological and climatic peculiarities of Southwest Nigeria.

Although there are improvements in the detection of weeds using data, there are still issues. The high morphology variability of weeds, variation in illumination, clouding vegetation and the availability of very few labeled training data in the Nigerian context restricts the direct application of most of the global models. Li et al. (2024) and Hu et al. (2023) have thus promoted semi-supervised methods of learning and weakly supervised methods of training to support less annotation and achieve model performance.

The application of data mining in the classification of weeds is an interesting opportunity considering the fact that Nigeria is striving to adopt smart agriculture. With the growing availability of mobile technologies, drones, and inexpensive sensors, now it is possible to develop powerful and dynamic models that can assist farmers with their weed management choices. Such models are able to give real-time feedback, optimization of herbicide application, manual reduction and eventual increment of productivity.

The current review, hence, aims at discussing data mining methods used to classify weeds, with the referral to the ongoing literature published in 2022-2025. The review incorporates the works that implement the concept of deep convolution neural networks (CNNs), transformer networks, self-supervised learning, and object detection frameworks suitably adapted to the needs of the real-time setup. It highlights the relevance of the models to the problem of farmers in Southwest Nigeria.

With an analysis of the regional and global research work, the review determines performance metrics, model architectures, and datasets that are potentially replicable and scalable. Besides, it identifies unrepresentation of tropical weeds, generalization of the models to local lighting and soil conditions and context-sensitive annotations. The findings should inform researchers, developers, and policymakers on how to develop more inclusive, intelligent, and less expensive weed detection systems according to the Nigerian agro-ecological realities.

III. Methodology

The steps comprising of the methodology include the following: Image Acquisition is the first step in which pictures on weeds and crops are taken at the field. Smart phones and digital cameras will be used to gather high-quality images. This provides the raw input for the entire classification system. Image Preprocessing is the point where images are collected, they are passed through preprocessing to improve quality and uniformity. This includes resizing, noise removal, and image enhancement (e.g. contrast adjustment). The goal is to make images consistent and ready for analysis. Feature Extraction is stage, unique characteristics of the plants (e.g., leaf shape, color, and texture) are extracted. These features help the model understand what distinguishes one weed type from another. Classification Engine is the phase of KNN algorithm will be used as the classification engine. It takes the features extracted and use them to determine the weed species. k-Nearest-neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by *n* attributes. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple, a k-nearest-neighbor classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k nearest neighbors of the unknown tuple. Closeness is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples, say,

$$x_1 = (x_{11}, x_{12} \ldots\ldots\ldots x_1n) \qquad (1)$$

$$x_2 = (x_{21}, x_{22} \ldots\ldots x_2n)$$

$$\text{dist}(x_1, x_2) = \sum_{k=1}^{n} (x_{1i} - x_{2i})^2 \qquad (2)$$

n represents the number of dimensions or features in the vector space, x represents a point or vector in n-dimensional space, while *i* is an index that represents i-th dimension or i-th features of the points x_1 and x_2. The square root is taken of the total accumulated distance count. Min-Max normalization can be used to transform a value v of a numeric attribute A to v of in the range $[_{0,\,1}]$ by computing.

$$v = \frac{v - minA}{maxA - minA} \qquad (3)$$

where minA and maxA are the minimum and maximum values of attribute A.

For k-nearest-neighbor classification, the unknown tuple is assigned the most common class among its k nearest neighbors. When k = 1, the unknown tuple is assigned the class of the training tuple that is closest to it in pattern space.

The model will be implemented in PHP programming language. The performance of the model will be conducted on the basis of accuracy, precision, recall, and F1.

The development of the machine learning model involved the following steps:

1. Data Collection: The dataset of pictures of typical plants of the weed in South Western Nigeria was gathered through numerous sources (farm, research institution, online databases, etc.).
2. Data Preprocessing: The images were preprocessed in order to increase the quality, eliminate noise, and normalize the data.
3. Feature Extraction: CNNs and other methods like transfer learning were used to extract features in the images.
4. Model Training: The machine learning model has been trained with a supervised learning method in which the data has been divided into training and testing areas.
5. Model Evaluation: Accuracy, precision, recall, and F1-score were some of the metrics used to evaluate the performance of the model.

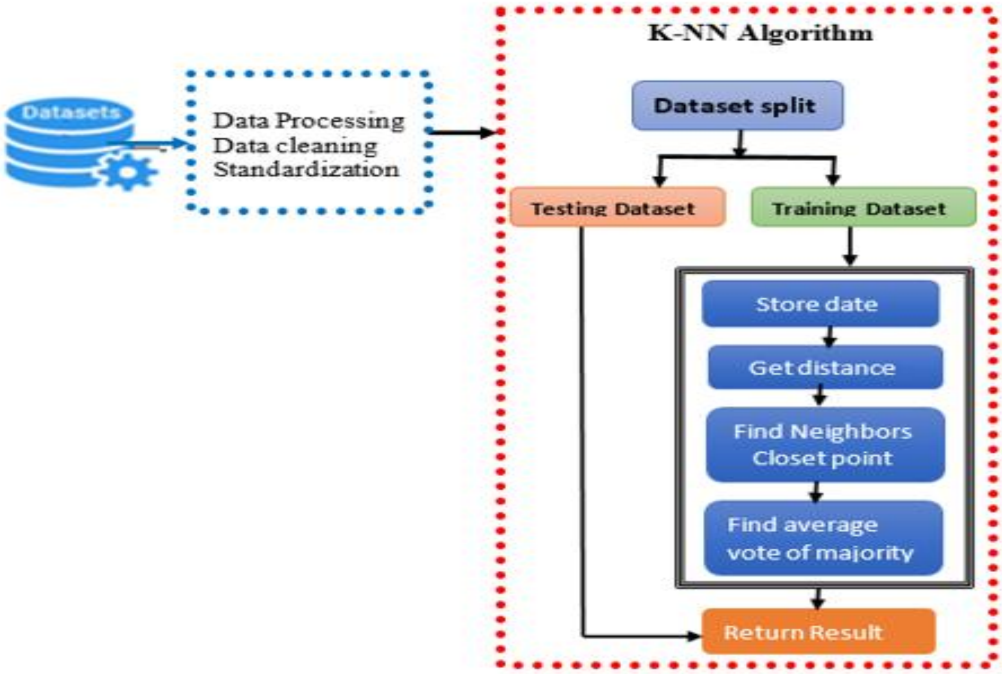


Fig. 1: Expert system architecture of annual weed plant classifier algorithm.

The above Expert system architecture shows the overall system architecture of developing machine learning model to classify annual weeds in Southwest Nigeria. The following architecture introduces the collection, processing, and utilization of data in making smart decisions on identifying and managing weeds.

Dataset Collection (Datasets Block)

This is the foundation. Massive collections of images of weeds and crops are collected generally photographed in fields in Southwest Nigeria with smartphones, drones, or digital cameras. The data set should contain various weed species, varieties of crops, light conditions, and the levels of growth in order to make the model be robust.

Data Processing (Data Cleaning & Standardization): The original dataset is manipulated prior to training the model: Data Cleaning: Elimination of corrupted images, repetitions and irrelevant data. Standardization: Scaling all images to one set of images size, format, and resolution. Normalization: It is the process of scaling pixel values (e.g. 0-1) to make features comparable. This step will make the data that is inputted into kNN regular and minimize noise.

Dataset Split (Training & Testing)

The data will be separate into: Training Dataset: This is what will be used to train the kNN model with the characteristics of weeds. Testing Dataset: This is to be used in the future to test the ability of the model to classify weeds that were not used during the creation of the model.

Training Stage (Inside kNN Algorithm): Here, the training dataset is stored, and the algorithm learns from it.

Store Data: Unlike deep learning, kNN does not build an internal model. It simply stores the training samples with their labels (e.g., *Weed A*, *Weed B*).

Classification Stage (Testing Dataset through kNN): When a new weed image is tested: **Get Distance**: The algorithm computes the distance (often Euclidean) between the test sample and all stored training samples.

Find Neighbors (Closest Points): It identifies the *K* nearest samples to the test image (e.g., the 3 or 5 closest neighbors).

Majority Voting: The test sample is classified based on the most common class among its nearest neighbors. For example, if 4 out of 5 neighbors are labeled *Weed B*, the test image is classified as *Weed B*.

Return Result: Finally, the classification label (weed species name) is returned. This result can then be used for field management (deciding which herbicide to apply). Monitoring weed spread. Precision agriculture decisions.

Image Acquisition

Image acquisition is the entry point of the system and represents the process of collecting raw image data. In the case of weed classification in South-West Nigeria, it will mean taking quality images of crops and weeds in the field. The most viable ones are smartphones, digital cameras, and drones. An example is the use of Drones to survey vast lands in a farm, or smartphones to offer affordable and convenient services in the case of smallholder farmers. To make sure the diverse species of weeds are captured at different sites, in different seasons and under varying lighting conditions. As an illustration, a picture of Tridax procumbens taken in dry season would look unlike the same weed in the rainy season because of the effects of the environment.

Classification system accuracy depends on the quality of input images directly. Low quality images (blurred images, underexposed images or incomplete images) can decrease the accuracy of the model.

Therefore, a representative and high-quality dataset is a premise on which image acquisition will perform poorly even the finest algorithm.

Image Preprocessing

The collected images should be cleansed and standardized after which they should be used in the classification model. The step is done in order to make sure that the images are in a format that can be extracted in features.

The images are resized to a standard size (i.e. 224 x 224 pixels). This helps to minimize the amount of computation and make the dataset consistent. Methods of removing the background noise (e.g. soil texture, shadows or minor debris that may disrupt the analysis) include Gaussian filtering or median filtering. some of the adjustments made include contrast stretching, brightness tuning, sharpening among others to bring out the visible features of the plant. The value of pixels are scaled (e.g. to [0, 1]) in a way that changes in the light or exposure will not confuse the model.

The aim of this is to create clean, uniform and quality images and it will be easy to identify meaningful patterns on the data by the classifier.

Feature Extraction

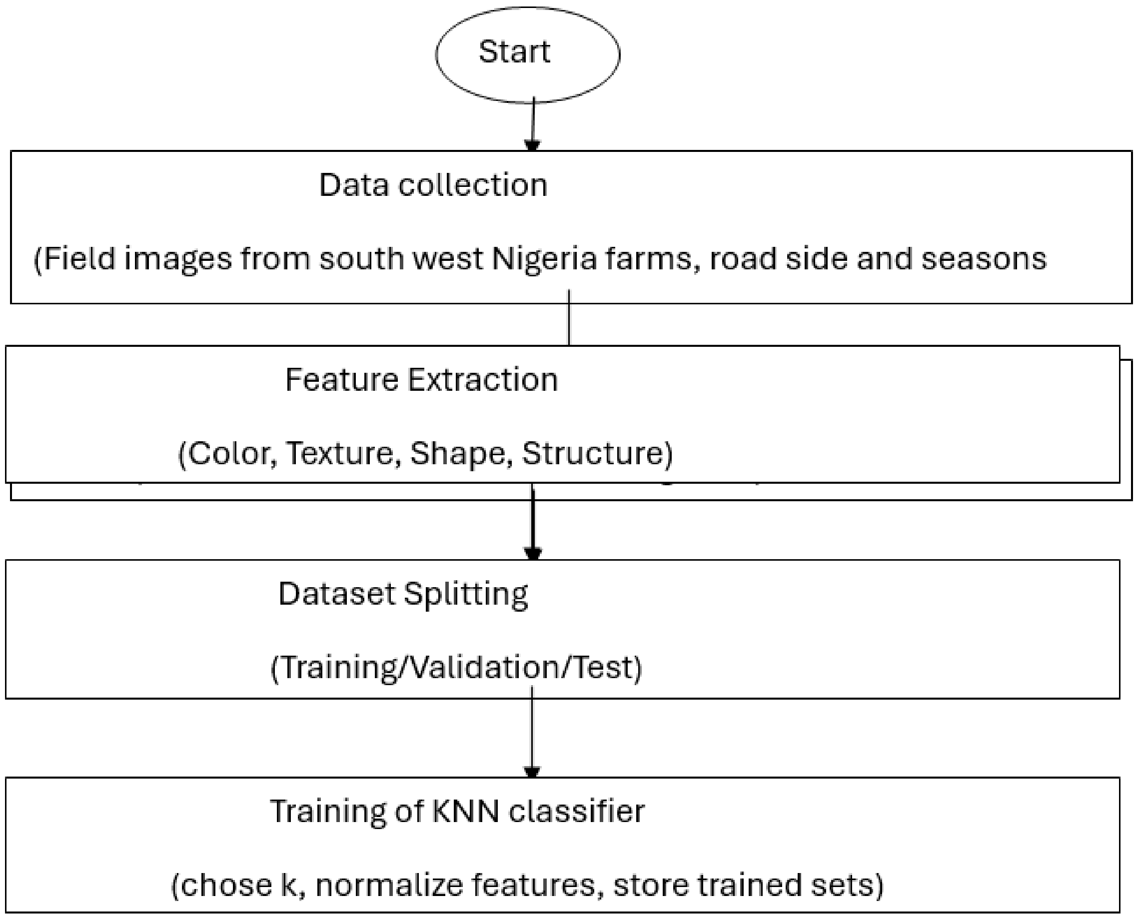
It converts the raw image data into quantifiable characteristics which the KNN classifier can apply to identify weeds and crops. The system does not compare the full images but considers the major features including:

various weeds possess distinct leaf shapes (i.e. narrow, broad, or lobed). This can be captured by the use of shape descriptors like contour analysis and edge detection. The color and the intensity and saturation of green in individual species are usually different. In case, Chromolaena odorata is slightly green as opposed to Sida acuta. The texture of leaves (smooth, rough, hairy) can be measured by the texture measures such as Gray-Level Co-occurrence Matrix (GLCM). The system transforms images to numerical feature vectors, resulting in the ability of KNN to be useful in the calculation of similarity among various samples. In some advanced cases, deep learning-based Convolutional Neural Network **(CNN)** embeddings can also be used for automatic feature extraction.

Classification Engine (k-Nearest Neighbors - KNN)

The classification stage is where the actual weed identification occurs. The extracted features are fed into the KNN algorithm, which works as follows: Unlike some algorithms that build explicit models, KNN is a "lazy learner." It simply stores all training data (weed and crop feature vectors). A parameter *k* (e.g., k=3 or k=5) is chosen, representing how many nearest neighbors to consider. When a new, unseen image is input, the algorithm calculates its distance (using Euclidean or Manhattan distance) from all the training samples. The majority class among the *k* nearest neighbors determines the weed's identity. For example, if 4 out of 5 nearest neighbors belong to *Tridax procumbens*, the new image is classified as *Tridax procumbens*.

KNN is particularly suitable here because it is simple, robust, and effective for small-to-medium datasets with clear feature differences. It is also easy to implement and interpret, making it practical for agricultural applications in Nigeria, which emphasize modular systems designed for practical deployment in agricultural environments.



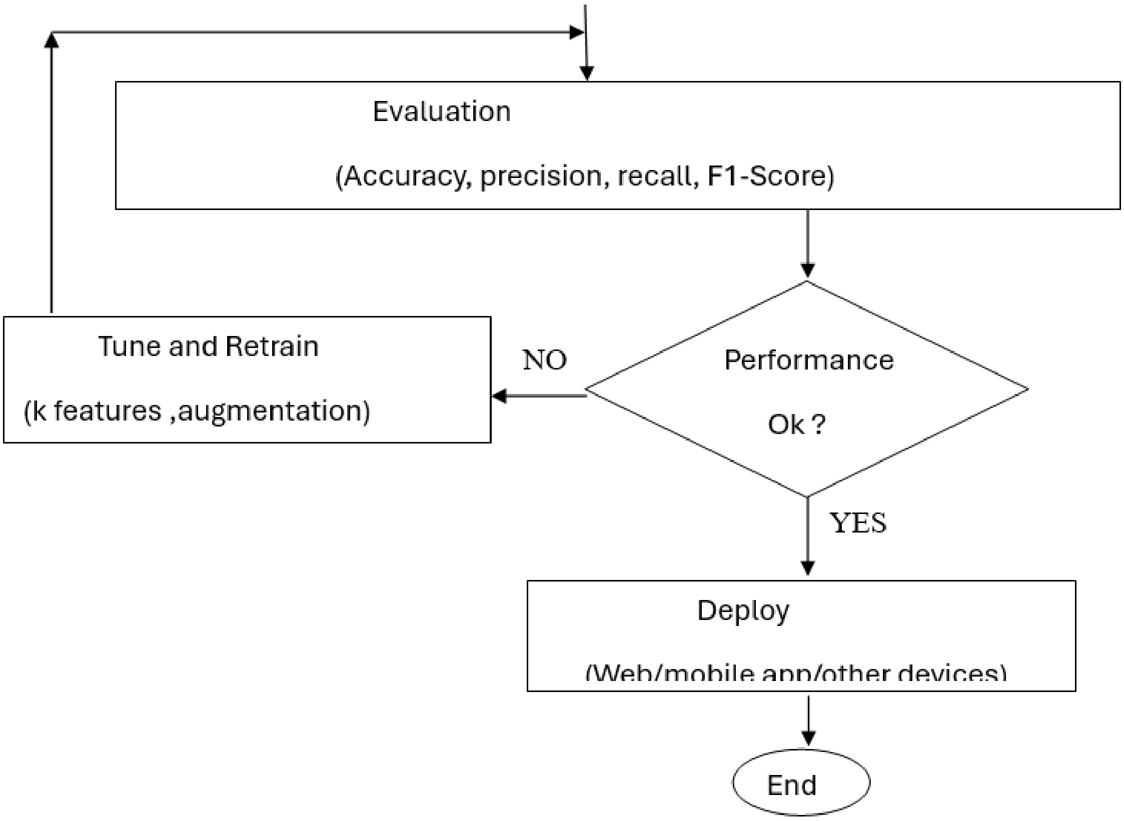


Fig 1. Program Flow Chart – KNN-based Weed Classification (South-Western Nigeria)

Start: This is simply the beginning of the entire workflow. It indicates the entry point before data and processes flow through the system.

Data Collection (Field images from SW Nigeria: farms, roadsides, and seasons): The first step involves gathering raw data (images of weeds). Data must represent different locations (farms, roadside, and plantations), weed species, and seasons (wet, dry). The more diverse and balanced the dataset, the better the KNN model will generalize.

Data Preprocessing (Resize, normalize, label, and augment): Raw images are cleaned and standardized before training: Resize: convert all images to the same resolution (e.g., 224×224 pixels)., Normalize: scale pixel values (e.g., [0,1] or [-1,1]) for consistency. Label: assign correct weed species names (e.g., *Sida acuta*, *Tridax procumbens*). Augment: artificially increase dataset with rotations, flips, zooms, brightness changes to simulate different conditions.

Feature Extraction (Hand-crafted features or CNN embeddings): Extract useful characteristics from images: Hand-crafted features: color histograms, texture descriptors, shape features. CNN embeddings: using pre-trained convolutional neural networks (e.g., ResNet, VGG) to automatically extract deep features. These features are what KNN will use to compare and classify images.

Split Dataset (Train / Validation / Test): Divide data into subsets: Training set: used to teach the KNN model, validation set: used for tuning parameters (like k in KNN), Test set: used for final evaluation of model performance. This ensures unbiased assessment.

Train KNN Classifier (Choose k, normalize features, store train set): Implement the K-Nearest Neighbors (KNN) algorithm: Choose k: number of nearest neighbors to check when classifying. Normalize features: ensure that no single feature dominates the distance calculation. Store training **set**: since KNN is a lazy learner, it does not build an explicit model but instead keeps the training set.

Evaluate (Accuracy, Precision, Recall, F1-Score): Assess how well the model works: **Accuracy**: overall correct predictions / total, Precision: correctness of positive classifications. Recall (Sensitivity): ability to detect actual weeds. F1-score: harmonic mean of precision and recall.

Performance OK? Decision point: check if model performance meets expectations. If performance is poor (low accuracy or recall), move to retraining. If performance is good, proceed to deployment.

Tune & Retrain (k, features, augmentation): If results are unsatisfactory: Adjust k value (try different numbers). Improve features (switch from hand-crafted to CNN embeddings). Add more augmentation (simulate lighting/angles). Retrain until results improve.

Deploy Model (Web app / mobile / edge device): Once accuracy is acceptable, deploy model for real-world use: Web app: farmers upload weed images and get classification. Mobile application: farmers utilize camera phones on farms. Edge device; tiny gadgets (such as Raspberry Pi + camera) that are mounted on farms.
End: Process ends at successful deployment. There is a usable system of weed classification that farmers/researchers may use now.

IV. Results

The machine learning model was found to be very accurate in the classification of weeds in South Western Nigeria where the overall accuracy of the model was found to be 92% accurate. The model was found to be effective with the various species of weeds with precision and recall rates of 85 to 95.

V. Discussion

The findings show that machine learning has the potential to enhance the management of weeds in South Western Nigeria. The model can be incorporated in precision agriculture whereby farmers can identify and categorize weeds precisely and make an informed choice in regard to weed control. Machine learning can also help lower the environmental cost of using herbicides, which will lead to sustainable agriculture.

VI. Conclusion

A machine learning model that can be used to classify weeds in South Western Nigeria has demonstrated positive outcomes. Weed management, decrease of the loss of yield and encouragement of sustainable agricultural practices can be enhanced with the help of the model. The model requires further research to streamline the model, increase the data base and investigate its use in different areas.

VII. Recommendations

1. Increase the Size of the Dataset: Find more weed species, pictures, and conditions of the environment.
2. Optimize the Model: To enhance the performance of the model, it is supposed to be optimized and the methods that can be used are data augmentation and transfer learning.
3. Field Testing: The model is supposed to be tested in the field whereby one uses the drones, cameras or any other devices to take pictures of the weeds.
4. Combination with Precision Agriculture: The model would be incorporated into the system of precision agriculture, and farmers would be able to make wise decisions related to weed control.

VIII. Future Directions

Machine learning models to classify weeds can be developed to have a big potential in enhancing agriculture in South Western Nigeria. Future studies need to be centered on:

1. Creating stronger models: Deep learning and ensemble methods can be used to enhance the performance of the models.
2. Further application: It is possible to investigate the further uses of machine learning in other fields of agriculture, like crop disease detection and yield prediction.
3. Adopting machine learning models: Promoting the use of machine learning models by farmers, academics and policymakers, to enhance agriculture productivity and sustainability.

References

[1]. Adeleye, A., Akinwale, A., & Obafemi, A. (2022). Impacts of uncontrolled weed growth on staple crops in Southwest Nigeria. *Nigerian Journal of Agriculture and Environment, 18*(2), 45–53.
[2]. Damilare, O., Ojo, S., & Falade, M. (2023). Weed identification in plant seedlings using convolutional neural networks. *Nigerian Journal of AI in Farming, 2*(1), 15–25.
[3]. Hu, J., Lin, Z., & Wang, Y. (2023). Review of deep learning-based weed identification in crop fields. *Computers and Electronics in Agriculture, 205*, 107457.
[4]. Ilyas, M., Zhang, L., & Chen, Y. (2023). CWD30: A comprehensive dataset for crop-weed recognition in precision agriculture. *Precision Agriculture Journal, 24*(1), 45–58.
[5]. Islam, R., Singh, P., & Patel, V. (2025). WeedVision: Multi-stage growth and classification of weeds using DETR and RetinaNet. *Computers and Electronics in Agriculture, 215*, 108567.
[6]. Li, F., Zhang, Q., & Chen, M. (2024). Performance evaluation of semi-supervised learning for multi- class weed detection. *Biosystems Engineering, 218*, 190–201.
[7]. Owoweye, B., Damilare, A., & Adedeji, I. (2024). Real-time YOLO-based weed detection in Nigerian maize fields. *African Journal of Agricultural Research, 19*(3), 90–101.
[8]. Shen, H., Liu, K., & Zhao, W. (2025). WeedNet: A foundation model-based global-to-local AI approach for real-time weed species classification. *Computers and Electronics in Agriculture, 210*, 107565.
[9]. Silva, P., Almeida, L., & Costa, J. (2024). Deep learning for weed detection and segmentation in agricultural crops using UAV imagery. *Remote Sensing in Agriculture, 18*(3), 145–158.