

DRONE-BASED CROP TYPE IDENTIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS: AN EVALUATION OF THE PERFORMANCE OF RESNET ARCHITECTURES

AJAYI, O.G^{1*} and OLUFADE, O. O²

¹Department of Land and Spatial Sciences, Namibia University of Science and Technology, Windhoek, Namibia,
oajayi@nust.na, ogbajayi@gmail.com

²Department of Surveying and Geoinformatics, Federal University of Technology, Minna, Nigeria,
oluwapelumiofufade@gmail.com

KEYWORDS: ResNet Architecture, Deep learning models, Convolutional Neural Network, Precision Agriculture, Smart farming, Drone mapping

ABSTRACT:

This study investigates the application of deep learning techniques, specifically ResNet architectures, to automate crop type identification using remotely sensed data collected by a DJI Mavic Air drone. The imagery was captured at an altitude of 30 meters, maintaining an average airspeed of 5 m/s, and ensuring a front and side overlap of 75% and 65%, respectively. The pre-flight planning and image acquisition was facilitated through the Drone Deploy platform, yielding a dataset consisting of 1488 aerial photographs covering the study area. These images possess an average ground sampling distance (GSD) of 22.2 millimetres. The dataset was meticulously labelled with "maize" and employed to train three distinct ResNet architectures, namely ResNet-50, ResNet-101, and ResNet-152. The evaluation of these models was based on accuracy and processing time. Notably, ResNet-50 emerged as the most proficient, achieving an accuracy rate of 82% with a precision score of 0.5 after just two hours of initial training, while ResNet-101 and ResNet-152 architectures achieved 27% and 24% accuracy, respectively. These outcomes underscore the potential of ResNet-50 architecture, even with a limited dataset, as a valuable tool for precise crop-type classification within the precision agriculture domain.

1. INTRODUCTION

The world's population is projected to surge to nine billion by 2050, leading to a substantial 70% consequential surge in the demand for agricultural production (Radoglou-Grammatikis et al., 2020). This unprecedented global population growth, coupled with evolving agricultural practices, exerts substantial pressure on the agricultural sector to enhance productivity in order to meet escalating food requirements. Nevertheless, challenges such as diminishing cultivable land, climate fluctuations, and water scarcity have significantly complicated the pursuit of this objective. Consequently, there is a growing imperative to explore alternative approaches to conventional agricultural practices, with precision agriculture (PA) emerging as a promising solution.

Precision agriculture (PA) involves the use of computer-based technologies, data acquisition, analysis, and storage systems to collect and analyse data that can be used to inform site-specific input applications (Ajayi et al., 2023). It is a crop and soil management system that uses computers to collect and analyse data. One important aspect of precision agriculture is crop identification which involves classifying and mapping vegetation.

In recent times, there has been a burgeoning interest in the application of deep learning (DL) techniques, specifically Convolutional Neural Networks (CNNs), in the domain of crop identification.

CNNs stand out as a subset of artificial neural networks renowned for their exceptional prowess in image recognition tasks. They have been used successfully in a variety of domains, including agriculture, to classify crop types using remotely sensed data (Schmedtmann and Campagnolo, 2015).

The goal of crop identification is to develop a classification model that can take remotely sensed images as input and output the crop class for each instance. This can help to improve the accuracy of crop type identification.

This study aims to address the limitations of traditional methods for crop identification and explore the potential of using drone-based technologies, specifically CNNs, for this task. Traditional methods, such as visual observation of fields, can be time-consuming and inaccurate, especially for large-scale or remote areas. Drone-based technologies can provide a more efficient and accurate way to collect data for crop identification. A more accurate and efficient approach to crop type mapping can be developed by leveraging UAV images and deep learning algorithms which could have a significant impact on the agricultural sector, helping farmers to improve their yields and profits.

*Corresponding author

The primary objective of this study is to evaluate the performance of selected DL algorithms, focusing on RESNET architectures, for automatic crop classification. One particular aspect of interest is to examine the viability of training these models with limited image datasets, as obtaining large labelled datasets for every crop type can be challenging and time-consuming. We seek to determine which model architecture exhibits optimal performance in crop type identification tasks by evaluating the performance of these DL models with limited data.

The results of this study will have practical implications for agricultural monitoring and decision-making processes. Our study will provide reliable information regarding the types of crops being cultivated. This information serves as a valuable input parameter for systems designed to evaluate crop health and monitor agricultural practices. In turn, these insights can guide farmers and stakeholders in making informed decisions related to resource allocation, crop management, and yield optimization.

Hence, this study delves into the application of CNNs for crop identification, leveraging remotely sensed data from a small farmland affiliated with the Federal University of Technology, Minna, Nigeria. Specifically, we conducted a comprehensive comparison of the performance of various ResNet architectures, namely ResNet-50, ResNet-101, and ResNet-152, to assess their effectiveness in accurately identifying maize crops. Our objective is to evaluate the performance of these deep learning models when tested with limited datasets, shedding light on their capabilities in scenarios with restricted data availability.

2. LITERATURE REVIEW

Remote sensing (RS) has been extensively utilized as a potent tool for various agriculture applications due to its capacity for rapid, precise, and dynamic data collection without physical contact and at a low cost (Chipman et al., 2015; Mulla, 2013; Eskandari et al., 2020). Unmanned Aerial Vehicles (UAVs) have rapidly emerged as the preferred technology for numerous precision agriculture (PA) applications, including crop state mapping (Laursen et al., 2017; Popescu et al., 2020), crop yield prediction (Zhou et al., 2017; Yang et al., 2019), and disease detection (Su et al., 2018).

Deep Learning (DL) methods, characterized by their deep neural networks with multiple layers of abstraction, have significantly advanced the state-of-the-art in various domains, including computer vision and natural language processing (Liu et al., 2019). In the context of precision agriculture and UAV data, DL has proven to be a powerful and reliable technique for applications such as weed identification (Canals et al., 2018), crop and plant counting (Geus et al., 2019), and land cover and crop type classification (Ferrari et al., 2021).

DL has demonstrated remarkable accuracy in precision agriculture, outperforming traditional Machine Learning (ML) methods. For instance, Bah et al. (2018) reported a performance gain of over 20% in weed detection using a DL model compared to ML methods. These successes

have motivated the increased adoption of DL in various PA applications.

Deep Learning (DL) is a subset of Artificial Neural Network (ANN) methods, characterized by its deep and complex neural network architectures (Hinton et al., 2006). Advances in hardware capabilities and the availability of large labelled datasets have enabled efficient DL training and inference, surpassing ML methods in numerous applications. In agricultural contexts, DL models are often based on Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and Convolutional Neural Networks (CNNs).

CNNs represent a specialized category within the realm of deep, feed-forward artificial neural networks, specifically tailored for computer vision applications in the field of machine learning (Schmidhuber, 2015; LeCun & Bengio, 1995). CNNs are ingeniously engineered to efficiently process grid-like data, such as images, rendering them particularly adept for supervised image processing and computer vision tasks. Their architectural composition typically encompasses three distinct hierarchical components: convolutional layers, pooling layers, and fully connected layers (Canziani et al., 2016). The convolutional layers serve the pivotal role of extracting essential features from input images. These features are then subjected to dimensionality reduction through the pooling layers, ultimately culminating in the fully connected layers, which act as classifiers, making decisions based on the learned features (Schmidhuber, 2015). Well-established CNN architectures include AlexNet (Hinton et al., 2012), GoogleNet (Liu et al., 2015), and ResNet (Ren et al., 2016).

One of the notable strengths of CNNs lies in their versatility regarding the types of input data they can accommodate. While they are renowned for their proficiency in image processing, CNNs are not constrained solely to visual data. They possess the capacity to process a wide spectrum of data types, including audio, video, speech, natural language, and more (Karpathy et al., 2014; Kim, 2014; Abdel-Hamid et al., 2014). This adaptability underscores their relevance in a broad array of applications beyond traditional computer vision tasks.

Crucially, CNNs distinguish themselves from traditional Artificial Neural Networks (ANNs) through their ability to tackle complex problems with remarkable speed, largely attributable to key architectural features such as weight sharing and the utilization of intricate models. These elements enable extensive parallelization of computational tasks (Pan & Yang, 2010). This characteristic empowers CNNs to expedite the learning process, making them particularly suited for large-scale issues where time-efficient solutions are paramount.

However, it is important to note that CNNs' effectiveness and accuracy are contingent on the availability of appropriately extensive datasets. The size of these datasets can vary significantly depending on the complexity of the problem under investigation. In scenarios where the subject matter is intricate and multifaceted, a larger dataset becomes indispensable to adequately characterize the nuances of the problem,

thereby enhancing the chances of successful classifications.

In recent years, the adoption of Deep Learning (DL) models in precision agriculture has gained substantial traction due to their potential to enhance both accuracy and speed within automated systems. DL techniques have found application across various facets of precision agriculture, encompassing tasks such as the detection and classification of plant diseases and pests, crop type classification, weed identification, plant growth monitoring, and plant health estimation. Several studies exemplify the versatility and effectiveness of DL models in these contexts.

For instance, Li et al. (2022) tackled the critical issue of identifying rice pests, which profoundly impacts rice crop yields, highlighting the limitations of existing rice pest datasets, such as small sample sizes and data imbalances. They created a substantial dataset, IP_RicePests, comprising 8,248 images across 14 categories, using web crawling and manual screening techniques. This dataset is further expanded to 14,000 images through ARGAN data augmentation. The authors employed transfer learning with pretrained VGGNet, ResNet, and MobileNet parameters from the ImageNet dataset for rice pest image classification. Experimental results revealed strong recognition accuracy across all three models, with VGG16 fine-tuning yielding the highest accuracy. Importantly, ARGAN data augmentation consistently improved the performance of the model. The study demonstrates the efficacy of CNNs combined with transfer learning and ARGAN data augmentation in mitigating sample size challenges, enhancing rice pest identification efficiency, and providing valuable support for the field.

In a similar vein, Yalcin et al. (2016) proposed a Convolutional Neural Network (CNN) architecture designed for the classification of diverse plant types based on image sequences. Their work entailed a rigorous comparison of the CNN approach against alternative methods, including Support Vector Machine (SVM) classifiers with various kernels and feature descriptors like Local Binary Patterns (LBP) and GIST. Impressively, the CNN model outperformed its counterparts, achieving an outstanding accuracy rate of 97.47%, notably surpassing the accuracy range of 74.92% to 89.94% exhibited by the alternative methods.

Furthermore, Ajayi et al. (2022) delved into the applicability of Artificial Neural Networks (ANN) for automatic crop type classification. Leveraging a straightforward feed-forward ANN for classification, they worked with diverse UAV datasets encompassing crops such as maize, groundnuts, soya, rice, beans, yam, and other non-crop features. Their results showcased the ANN's strong performance, boasting an overall training accuracy of 87.7%. Subsequent analysis, including the computation of the confusion matrix, substantiated an impressive overall accuracy of 0.9393.

In another innovative approach, Ajayi & Ashi (2023) explored automatic weed identification and classification. Their study centered on the effects of varying training epochs on the accuracy of a Faster Region Convolutional Neural Network (RCNN) model. Over the course of five

different training epochs, with no predefined intervals (10,000, 20,000, 100,000, 200,000, and 242,000), they observed a notable trend. The model's performance consistently improved with each increment in epoch count, though it eventually reached a saturation point at the 242,000 epoch mark.

These studies collectively underscore the robust applicability of DL models in precision agriculture, exemplifying their capacity to excel in diverse applications, from crop type classification to disease detection and weed identification.

ResNet stands as a pivotal milestone in the evolutionary race of CNN architectures. Its transformative impact lies in its ground breaking concept of residual learning within CNNs and its efficient methodology for training exceptionally deep networks. The ResNet family comprises several variants, with ResNet 50, ResNet 101, and ResNet 152 differing primarily in the number of layers they encompass. To be precise, ResNet 50 boasts 50 layers, ResNet 101 extends to 101 layers, while ResNet 152 reaches a substantial depth of 152 layers.

The empirical findings of He et al. (2016) underscore the superiority of ResNet architectures over their predecessors. Through rigorous experimentation, they demonstrated that ResNet models with 50, 101, or 152 layers exhibit significantly lower errors in image classification tasks compared to plain networks with 34 layers. ResNet also achieved a notable milestone by outperforming the widely recognized COCO image recognition benchmark dataset by a substantial margin of 28%, as reported by Lin et al. (2014). This achievement underscored the critical importance of representational depth in numerous visual recognition tasks, solidifying ResNet's reputation as a game-changer in the field.

At the core of ResNet's innovation lies the concept of bypass pathways, also known as skip connections, which draws inspiration from Highway Networks. These pathways address challenges encountered during the training of deep networks, particularly the issue of vanishing gradients. In contrast to Highway Network gates, ResNet's skip connections are data-independent and parameter-free. They play a pivotal role in preserving the flow of residual information throughout the layers, ensuring that identity shortcuts remain open. This design choice is in stark contrast to Highway Networks, where gated shortcuts can be closed, rendering the layers to represent non-residual functions. The inclusion of residual links, or shortcut connections, in ResNet architectures significantly accelerates the convergence of deep networks. By facilitating the smooth passage of gradient information, ResNet effectively mitigates the problem of gradient diminishing, a common challenge in training deep neural networks.

3. MATERIALS AND METHODS

Figure 1 illustrates the architectural workflow employed in designing the crop identification system utilizing ResNet models. The key stages of this workflow are succinctly delineated as follows:

a) **Data Collection:** The process initiates with the acquisition of a dataset, accomplished through the

deployment of a well-equipped Unmanned Aerial Vehicle (UAV) over a designated small farmland.

b) **Data Preparation:** Subsequently, the acquired UAV data undergoes pre-processing and processing steps to refine and optimize its quality and usability.

c) **ResNet Model Development:** In this phase, a dedicated ResNet algorithm is designed and developed, tailored specifically for the task of crop identification.

d) **Data Training and Testing:** The dataset is then subjected to data processing, involving the training and testing of images utilizing ResNet-50, ResNet-101, and ResNet-152 models, a crucial step in the workflow.

e) **Performance Evaluation:** Finally, an assessment of the performance of each ResNet architecture is conducted, allowing for a comprehensive understanding of their effectiveness in the context of crop identification.

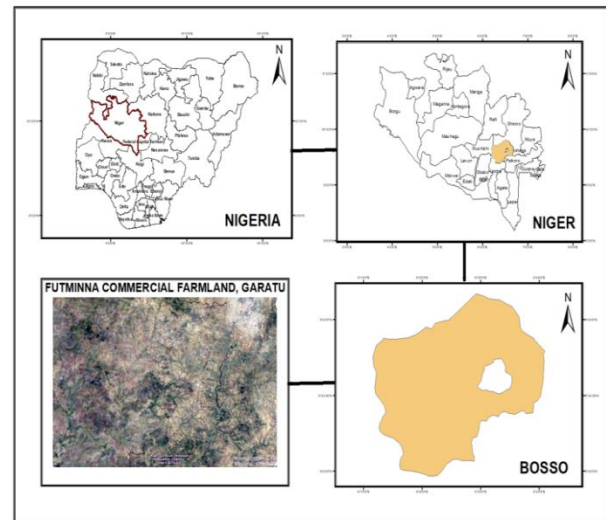


Figure 2. The study area



Figure 1. Flow diagram of the development and implementation of the ResNet architectures

3.1 Study Site

A portion of the Federal University of Technology Minna's commercial farmland was used for this study. The farm is located at Gidan-Kwano near Garatu in Minna, Niger State and it covers about 21 hectares. It is located between 9°32'3.7968" N and 6°25'22.411" E, as well as 9° 27'52.506" N and 6°27'13.957" E (see figure 2). This small farmland predominantly contained maize crops (*Zea mays*).

3.2 Data Acquisition

Approximately 1488 images of the designated study area were obtained using a DJI Mavic Air drone. The drone was operated at an elevation of 30 meters and maintained an average airspeed of 5 meters per second, while ensuring a front overlap of 75% and a side overlap of 65%.

3.3 Image pre-processing

The preprocessing procedures executed on the obtained images are elaborated upon below and are also visually depicted in Figure 3;

- Resizing:** Resizing is a crucial step that standardizes the size of all images to a fixed size. This operation helps to reduce the computational load and enables faster training. Hence, the images were resized to 224x224.
- Normalization:** The process of image normalization entails adjusting the pixel values to a range that aligns with the requirements of the ResNet architecture. Typically, this entails rescaling the pixel values to fall within the range of either 0 to 1 or -1 to 1. Normalization mitigates the influence of lighting variations and enhances the overall accuracy of the model. In this particular case, the images were normalized to span the range from 0 to 1.
- Image cropping:** This process submaps a smaller region of interest from a larger image. It helps to focus the model on the most important parts of the image and improves model accuracy.
- Splitting data:** The entire dataset which consists of 1488 images was split into two subsets for training and validation which was 70% and 30% respectively.

3.4 Training Process and Performance Evaluation

The model training process was conducted utilizing Kaggle GPU 100, a cloud-based system made available through the collaborative efforts of NVIDIA and Kaggle. A Python 3 programming notebook was employed in

conjunction with Jupyter notebook, and it was uploaded to Kaggle to execute the Python-based code responsible for constructing the ResNet models. In the course of training the ResNet model, a range of packages and libraries were integrated into the Python environment (as depicted in Figure 3).

Notably, for compatibility and various mathematical computation requirements, Numpy version 1.19.5 and TensorFlow version 1.15.2 were installed on the virtual machine. Keras was imported as a high-level API for building and training the ResNet architectures. Matplotlib was imported for visualizing the training and evaluation metrics of the three models while Pillow was imported for loading and preprocessing the image data before training the ResNet models. The dataset was imported into the Kaggle environment and an image path was created for storing the labeled images. The label name was written in a 'config file' defining the (maize field and crop field). Subsequently, the dataset underwent division into two subsets: a training set encompassing 70% of the data and a validation set comprising 30%. The images were uniformly resized to dimensions of 224 x 224, and a batch size of 32 was adopted. Furthermore, the images were transformed into boolean arrays across all indexes. To monitor the training's performance, a TensorBoard was installed and initialized. The training procedure was then initiated for the ResNet-50 model, and this process was subsequently repeated with minor adjustments for the other two ResNet architectures.

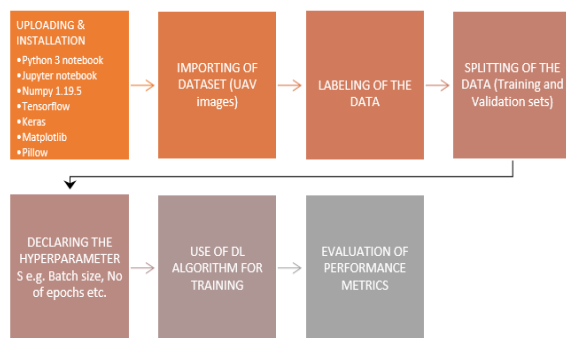


Figure 3. The process flowchart for model implementation

The evaluation of the ResNet architecture's performance was carried out using a set of critical metrics, including accuracy, precision, validation accuracy, and validation loss. Accuracy serves as a pivotal metric, demonstrating how effectively the model consistently makes correct predictions. A high accuracy score indicates that the classifier makes minimal errors in its predictions, while a low accuracy score suggests a higher frequency of prediction mistakes. Precision, on the other hand, delves into the quality of positive predictions made by the classifier. It measures the proportion of the classifier's positive predictions that are indeed accurate. A high precision value signifies that the classifier produces relatively few false positives, indicating its proficiency in making correct positive predictions with confidence. Conversely, a low precision score implies a higher likelihood of generating false positive predictions. Validation accuracy plays a crucial role during the training process by serving as a monitoring tool. It helps

in identifying the occurrence of overfitting, a phenomenon where a model excels on the training dataset but falters when presented with new, unseen data. Calculated after each training session, validation accuracy assists in selecting the best model among various architectures, hyper-parameters, or checkpoints. This metric provides valuable insights into the model's generalization performance, aiding in model selection decisions. Validation loss is another vital metric used in the evaluation process. It is computed by comparing the model's predictions to the actual values for each sample within the validation dataset. The validation loss is then determined by averaging the individual loss values across all samples in the validation dataset. This metric serves as an important indicator of how well the model is performing and aids in the fine-tuning of model parameters to achieve optimal results. The validation loss will be greater than the training loss when a model is overfitting, indicating that the model is not generalizing well to new data. Detailed information on these performance evaluation metrics including their mathematical formulation can be found in Ajayi and Ashi (2023).

4. RESULTS AND DISCUSSION

Table 1 presents the result of the performance evaluation obtained from the three ResNet architectures implemented in this study. The result shows that ResNet 50 outperformed ResNet101 and ResNet152 with an overall accuracy of 82%, precision of 50%, and validation accuracy of 51%.

Conversely, the ResNet-152 model displayed the least accuracy in classification, achieving an accuracy rate of only 24% and a precision of 30%, with a validation accuracy of 27%. For a visual representation of the training progress and performance of the three ResNet architectures, Figures 4-6 provide graphical depictions of validation accuracy and loss.

In these graphical representations, the blue line corresponds to the training data, while the orange line represents the validation data. The vertical axis (x-axis) reflects the accuracy values, while the horizontal axis (y-axis) indicates the number of training epochs. These graphs effectively illustrate the interplay between training and validation accuracy, as well as training and validation loss. The outcomes from these graphs underscore the ResNet-50 model's resilience in the context of automatic crop classification, even when dealing with limited datasets.

Metric	ResNet 50	ResNet101	ResNet152
Accuracy	82%	27%	24%
Precision	0.5	0.4	0.3
Val accuracy	51%	31%	27%
Val loss	20%	25%	20%

Table 1. Performance of the three ResNet architectures

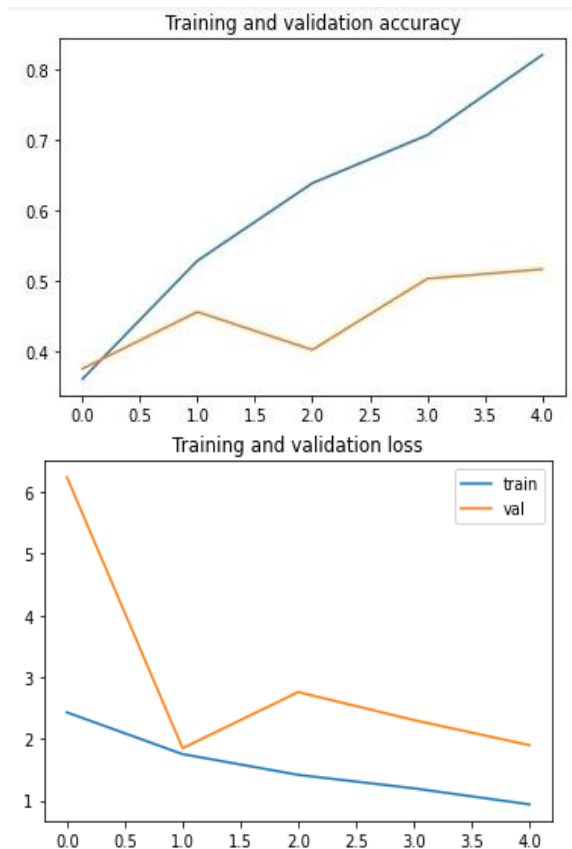


Figure 4. The accuracy and loss graphs for training for ResNet50

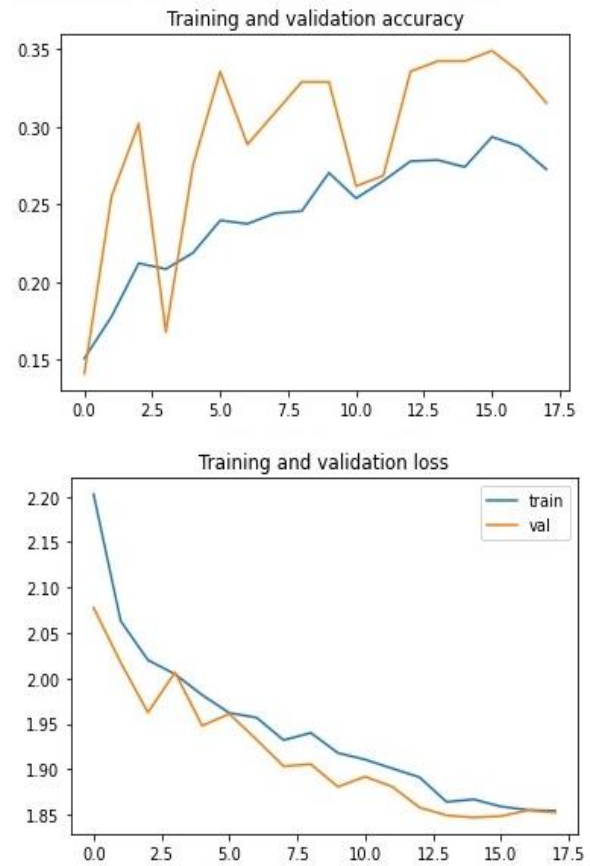


Figure 5. The accuracy and loss graphs for training ResNet101

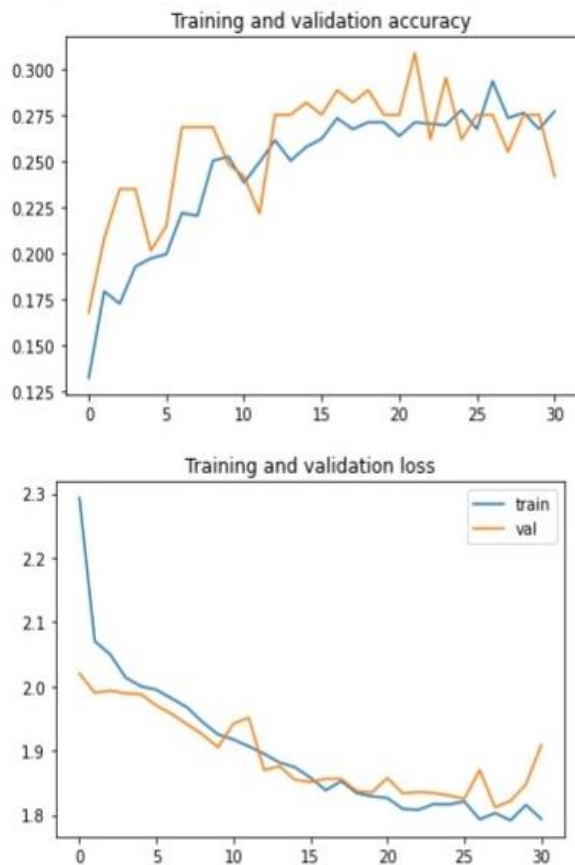


Figure 6. The accuracy and loss graphs for training ResNet152

Figure 4 which shows the accuracy and loss graphs for training ResNet50 depicts that the training accuracy of ResNet 50 is higher than its validation accuracy i.e. the level at which the ResNet 50 detects images is higher than the rate of how many images it can predict. However, the validation loss of ResNet50 is greater than the training loss which shows that the model is overfitting, indicating that the model is not generalizing well on the UAV dataset. On the other hand, the accuracy and loss graphs for training ResNet101 (Figure 5) show that the training accuracy of ResNet 101 is lower than the validation accuracy, however, the validation loss of ResNet 101 is lower than the training loss and this shows that the model is fitting well, indicating that it is generalizing well with UAV dataset. Finally, from Figure 6, it was observed that the training accuracy of ResNet 152 is interwoven with the validation accuracy which is the level at which the ResNet 152 detect image depending on the rate of how many images it can predict. However, ResNet 152 validation loss is interwoven with its training loss and this shows that the model is underfitting but also indicates that the model would generalize well with the UAV dataset if the volume of the dataset is increased significantly.

5. CONCLUSION

In conclusion, this research aimed to assess the suitability of ResNet architecture for maize crop classification using limited drone data. The classification results indicate that, despite the constraints of limited datasets, ResNet-

50 emerged as a more reliable choice for precision agriculture and crop identification compared to ResNet-101 and ResNet-152, which demonstrated a higher dependency on larger datasets for accurate performance. Notably, this study highlights the feasibility of utilizing UAV data, which is cost-effective and widely accessible, for crop type detection on small agricultural farms.

The achieved accuracy of 82% in crop type classification using the ResNet-50 model underscores the potential of deep learning models in accurately evaluating crop types. However, it is important to note that the performance of ResNet-101 and ResNet-152 could have been further improved with a substantially increased dataset, suggesting their potential for higher accuracy given more extensive training data.

This study contributes valuable insights to the field of agricultural research by showcasing the applicability of deep learning models, specifically ResNet architectures, for crop identification tasks using limited drone data. By emphasizing the effectiveness of UAV data in achieving satisfactory results, this research highlights the potential for cost-effective and efficient monitoring of crop types on small-scale agricultural farms.

In future studies, it is recommended to explore the performance of ResNet-101 and ResNet-152 with larger datasets to fully assess their capabilities in crop identification. Additionally, investigating the integration of additional processing steps, such as analyzing crop development stages and assessing crop health, could further enhance the practical applications of this research in improving agricultural practices and decision-making processes.

REFERENCES

- Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. 2014. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10), 1533-1545.
- Ajayi O. G., Opaluwa Y. D., Ashi J and Zikirullahi W. M. 2022. Applicability of Artificial Neural Network for Automatic Crop Type Classification on UAV-Based Images. *Environmental Technology and Science Journal*. 13(1), DOI: <https://dx.doi.org/10.4314/etsj.v13i1.5>.
- Ajayi O. G and J. Ashi. 2023. Effects of Varying Training Epochs of a Faster Region-Based Convolutional Neural Network on the Accuracy of an Automatic Weed Classification Scheme. *Smart Agricultural Technology* 3, <https://doi.org/10.1016/j.atech.2022.100128>
- Bah, M.D.; Hafiane, A.; Canals, R. 2018. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. *Remote Sens.*, 10, 1690.
- Canziani, A., Paszke, A., & Culurciello, E. 2016. An Analysis of Deep Neural Network Models for Practical Applications. *arXiv*, 1605(07678)
- Eskandari, R., Mahdianpari, M., Mohammad manesh, F., Salehi, B., Brisco, B., Homayouni, S. 2020. Meta-Analysis of Unmanned Aerial Vehicle (UAV) Imagery

- for Agro-Environmental Monitoring Using Machine Learning and Statistical Models. *Remote Sens.*, 12, 3511.
- He, K., Zhang, X., Ren, S., Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June pp. 770–778.
- Hinton, G.E., Salakhutdinov, R.R. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313, 504–507.
- Yalcin, H and S. Razavi, 2016. Plant classification using convolutional neural networks. *Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Tianjin, China, pp. 1-5, doi: 10.1109/Agro-Geoinformatics.2016.7577698.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. 2014. Large-scale video classification with convolutional neural networks. *Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1725-1732. Columbus, OH, USA.
- Kitano, B.T.; Mendes, C.C.T.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. 2019. Corn plant counting using deep learning and UAV images. *IEEE Geosci. Remote. Sens. Lett.*, 1–5.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. 2012. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.
- LeCun, Y., & Bengio, Y. 1995. Convolutional networks for images, speech, and time series. *The Handbook of brain theory and neural networks*, 3361(10).
- Li, Z., Jiang, X., Jia, X., Duan, X., Wang, Y., & Mu, J. 2022. Classification Method of Significant Rice Pests Based on Deep Learning. *Agronomy*, 12(9), 2096. <https://doi.org/10.3390/agronomy12092096>
- Lin Tsung-Yi, Maire Michael, Belongie Serge, Hays James, Perona Pietro, Ramanan Deva, Dollár Piotr, and Zitnick C Lawrence. 2014. Microsoft COCO: Common objects in context. *ArXiv*. /abs/1405.0312. <https://doi.org/10.48550/arXiv.1405.0312>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. 2019. Johnarson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.*, 152, 166–177.
- Mulla, D. J. 2013. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
- Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P.P., Garcia-Herranz, M., Kim, D.-H. 2021. Crop type mapping by using transfer learning. *Int. J. Appl. Earth Obs. Geoinf.*, 98, 102313.
- Pan, S.J & Yang, Q. 2010. A survey on transfer learning. *IEEE Transaction on Knowledge and Data Engineering*, 22, 1345-1359.
- Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., & Moscholios, I. 2020. A compilation of UAV applications for precision agriculture. *Computer Networks*, 172, 107148.
- Schmedtmann, J. and M. L. Campagnolo. 2015. Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control. *Remote Sens.*, 7, 9325-9346; doi:10.3390/rs70709325
- Schmidhuber, J. 2015. Deep learning in neural networks: An overview. *Neural networks*, 61, 85- 117
- Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., Li, Q., Guo, L., Chen, W.-H. 2018. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Comput. Electron. Agric.*, 155, 157–166.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- Yang, Q., Shi, L., Han, J., Zha, Y., Zhu, P. 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.*, 235, 142–153.
- Zhou, X., Zheng, H., Xu, X., He, J., Ge, X., Yao, X., Cheng, T., Zhu, Y., Cao, W., Tian, Y. 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.*, 130, 246–255.