Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 15s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Advancements in Weed Detection with Deep Learning Models and Image Augmentation Techniques

T. Maneesha[1*], B. Eswar Adithya[2], Dr. V. Chandra Prakash[3], T. Saketh[4], Dr. M Madhusudhana Subramanyam[5]

[1*235]Koneru Lakshmaiah Education Foundation, Veddeswaram, Guntur, Andhra Pradesh, India.

Email ID: maneeshathotakura25@gmail.com, eswaradithya78@gmail.com, vchandrap@kluniversity.in, mmsnaidu@yahoo.com

[4]Broadcom, Wipro gate, Electronic City phase 1, Bengaluru, Karnataka, India 560100.

Email ID: sakethchowdary321@gmail.com

Cite this paper as: T. Maneesha, B. Eswar Adithya, Dr. V. Chandra Prakash, T. Saketh, Dr. M Madhusudhana Subramanyam, (2025) Advancements in Weed Detection with Deep Learning Models and Image Augmentation Techniques. *Journal of Neonatal Surgery*, 14 (15s), 2221-2230.

ABSTRACT

Deep-Learning is one of the potent technique for automating weed detection procedures. This study offers a thorough investigation of deep learning methods for weed identification, emphasizing training-models such as Convolutional-Neural Networks (CNN), Autoencoders & Vision-Transformers. Specifically, CNNs are skilled at identifying hierarchical complex features from high dimensional images, while Autoencoders facilitate unsupervised feature learning, and Vision transformers which uses the power Attention Neural Networks enable selective focus on relevant regions in images. We review recent developments, difficulties, and potential paths in utilizing these deep learning models for weed detection, highlighting how they could transform farming methods by facilitating accurate and timely weed management strategies. Dataset contain images of carrot and weed which contain 250 training images out of which 130 carrot plant images and 120 weed image. There is 70% testing images with RGB range of 0 to 255. As we have only small dataset to ensure the model's resilience various argumentation where applied like flipping – horizontal/vertical, Magnified Range, brightness, height shift range and width shift range.

Keywords: *Deep learning, Computer vision, Tensorflow, Keras, ReLU, Stochastic Gradient Descent, vgg, transformers.*

1. INTRODUCTION

Training multi-layered artificial neural networks (thus the word "deep") to learn and predict from data is the basic terminology in deep learning, which is a subset of Artificial Intelligance & machine learning. Neural networks are modeled using the human brain structure and operations, this networks are made up of number of interconnected layers of nodes or neurons. Here is a brief overview of several important deep learning concepts. In that network, one neurons output serves as the input for another. In actuality, the procedure is very similar. After passing information through an activation function for introducing non-linearity, each neuron generates its output. Layers make up the structure of neurons. An input layer, one or more hidden layers, and an output layer make up the standard structure.

There are several kinds of deep learning architectures intended for specific tasks. Some common architectures: Feedforward Neural Networks (FNN): Just think of how all data can only go in one direction: from inputs to outputs. Convolutional Neural Networks (CNN): Generally, neural networks may be used in cases in which the input data occur in some kind of grid structure, such as when processing images. Here, the use of convolutional layers allows for adaptive learning of spatial hierarchies of features. Recurrent Neural Networks: Ideal to use whenever there is a form of sequential data: text, time series data, etc. RNNs contain connections forming directed cycles that allow them to display dynamic temporal behavior. Long-Short-Term-Memory Networks (LSTM): Those designed architectures of RNNs capable of learning and understanding the long-term dependencies in the sequential data.

Procedure for Training These deep learning models are taught using a technique known as backpropagation, which iteratively updates the model weights to reduce the discrepancy between the corresponding actual targets in a training dataset and the anticipated outputs. Activation Functions: To help the network discover intricate patterns in the data, the activation functions give the model non-linearity. Sigmoid, tanh, ReLU (Rectified Linear Unit), softmax, and others are a few frequently used

activation functions. Functions of Loss: The discrepancy between the goal values and the outputs produced by a model is known as the loss function. Loss function: In deep learning, the selection of a loss function is contingent upon the job being carried out, which may include classification or regression. Classification requires a cross-entropy loss function, and regression has mean squared error.

Deep learning models have a few hyperparameters that would be necessary to have tuned in order for the optimal performance of the model. It comprises, among other things, the network's learning rate, batch size, and number of layers or neurons. Experimentation and optimization techniques are frequently used for hyperparameter adjustment. Deep learning is used in a wide range of fields, such as recommendation systems, speech recognition, natural language processing, computers and vision, and more. It brought about amazing advances in fields like machine translation, object identification, picture categorization, and possibly even the emergence of driverless cars. The field of artificial intelligence and computer science known as computer vision seeks to give computers the capacity to understand and interpret visual data from the outside world. It is an aspect of computer technique in which techniques and algorithms are developed so that a computer can capture digital images or videos; process them; analyze, decode, or extract meaningful information; and copy the vision of a human eye.

Computer vision has been applied in all possible applications: from medical to military to financial and far beyond. One among the most exceptional areas is the application in autonomous vehicles, where computer vision assists with navigation, object detection, and comprehension of a scene while ensuring safe and efficient travel. In healthcare, it plays a very essential role as medical image analysis could be carried out to enable in the detection of diseases and even during surgery being performed by surgeons. Surveillance and security benefit from using computer vision where monitoring or the identification of suspicious activities or objects helps augment public safety. Augmented and virtual reality experiences are richly enhanced through the use of computer vision due to its being an overlay of digital content onto the real world to create environments that are immersive. In robotics, computer vision factors in object manipulation and navigation, human-robot interaction, which is areas driving automation; for agriculture, it monitors crops, predicts yield, detects pests for high efficiency and yield. Then, in retail, computer vision utilizes product recognition, inventory management, analysis of customer behavior to streamline work and enhance the services offered to customers. Despite these successes, computer vision faces an array of problems. They include managing variability in lighting, viewpoint, scale, occlusion, and cluttered backgrounds. Beyond the strong efficiency, interpretability, and fairness of computer vision systems, there is focus on research in progress.

2. LITERATURE REVIEW:

J. Sangeetha et al., 2018 [1], this work addresses the problem of agricultural wastes' management through composting, where it aims to stabilize, sanitize, and make a significant amount of waste environmentally friendly, and hence, improve soil fertility. While methods like Carbon to Nitrogen ratio (C:N) or Germination Index (GI) are widely utilized in determining the quality of compost, they require ample time and complexity. With regards to this, the authors find applicability in using the Faster R-CNN, which captures the image characteristics at multilevels of cascaded layers of convolution with activation functions. From the study of images of various composting stages, Faster R-CNN would quickly assess the development of compost, thus becoming a more practical and effective means of assessing the quality of compost, and destined to revolutionize waste management practices of agriculture drastically.

Nanyang Zhu et al 2018 [2], Brief overview of all major DL algorithms specially for agricultural researchers, who tend to apply DL techniques rather without proper insight into their principles. This has made the paper delve into the ideas, restrictions, application, training procedures, and DL algorithm sample codes, including CNN, RNN, and GAN. At the same time, the paper makes an analysis of existing DL applications in the agricultural sector to unveil emerging trends, challenges, and prospects in this field. By providing this holistic view, the authors in this work shall enable agricultural researchers in the best use of DL so that data analysis shall be enhanced and agricultural research progresses while praxis DL application in agriculture shall be improved.

Muhammad Hammad Saleem et al 2019 [3], The Transition of Plant Disease Identification and Categorization from Conventional Machine Learning to Modern Deep Learning Methods: Towards Increased Precision and Opportunities Saleem, Muhammad Hammad et al. The Development of Plant Disease Identification and Categorization from Traditional Machine Learning to the Contemporary Deep Learning Techniques: Towards Improved Accuracy and Possibilities. The paper proposes providing an overview of these DL models and their implementations based on the requirement for performance metrics for the evaluation process. Moreover, it detects gaps in current research that can further enhance transparency regarding the detection of diseases even before symptoms begin to appear. Ultimately, this paper aims at contributing to the development of DL-based approaches for plant disease detection as well as their classification in the final step.

S. Kavitha et al 2023 [4], focusing on real time fast accurate identification of medicinal plants in India through the vision based smart approach with DL models. Since medicinal plants are the very root of Indian culture and healthcare so, proper identification requires a perfect identification which is quite tedious. To ease this process, the research article focuses on six specific herbs and collects 500 images for each one of them, resized and augmented to enhance the dataset. Using the

MobileNet DL model, the research manages to attain an exceptional accuracy rate of 98.3% to identify medicinal leaves. Then, the DL model is uploaded into the cloud, thus making it possible to build a mobile app for identification of leaves in real-time. With the new approach taxonomical gaps are filled with great interest in both botany and machine vision, promising very efficient prospects for the identification of plants in the coming future.

Jun Liu et al 2021 [5], concentrating on the use of digital image processing and deep learning technology to identify plant diseases and pests. The study emphasizes that deep learning outperforms conventional methods in this domain, acknowledging that pests and diseases are significant causes of reduced plant output. Lastly, it offers a review and prediction of future developments about the factors that motivate deep learning for the identification of plant diseases and pests. All things considered, this work offers important new perspectives on the developments and difficulties of applying deep learning to the detection of pests and plant diseases.

Alvaro Fuentes et al 2020 [6], In the aspect of implementing Deep-Learning in agriculture which aims at using non-destructive methods towards identification of plant diseases. "In real-worl scenarios, performance is relatively poor. Nonetheless, remarkable improvements have been reported". The proposed methods were tested on tomato plant disease, which was also gathered by the authors themselves along with providing annotations. Qualitative as well as quantitative results demonstrate the successful identification of plant diseases even in complex real-world scenarios. Some general insights from this research better illuminate how to design and apply methods for recognizing plant diseases with Deep Learning, highlighting its strengths and weaknesses.

Sunil G C et al 2022 [7], utilizing RGB-image texture elements to classify crop species & weeds in precision agriculture. SVM-based, VisualGroupGeometry 16 (VGG16) classification type of models are contrasted. The greenhouse environment yielded 3792 RGB photos in total, of which 1521 were of crops and 2271 were of weeds. The Relief feature selection technique has been used to determine the important features for constructing the prediction models. Four weed species were separated from six crop species using SVM and VGG16 classifiers. This study's shows the findings for the application of deeplearning algorithms—particularly VGG16—in weed detection for precision in agriculture site-specific controlled weeds.

Senthil Kumar Swami Durai a et al 2021 [8] Precise farming techniques have emerged, which play the most important role in the change of traditional farming norms. Using IoT, Data Mining, Machine Learning and other cutting-edge technologies precise farming tries to cater to the issues in the classic ways of farming. It helps farmers through much data collection, analysis and prediction to make the right decisions for crop growing, pest killing, and resource utilization. This proactive approach, in addition to increasing productivity, also minimizes the effects of negative factors arising from environmental uncertainty, such as erratic weather patterns and soil erosion.

Niranjan C Kundur et al 2022 [9], addresses the need for systematic improvement in crop productivity and quality in agriculture through the adaptation of efficient pest detection. Focused on the prompt and accurate identification of insect pests, this study offers a sophisticated approach tailored for wide-scale application. Utilizing deep learning algorithms that are particularly trained on IP102 data set including most of 75,000 images, the system thus allows for real-time detection of pests of significant agricultural impact. The use of MATLAB environment combined with the K-Means clustering algorithm enhances the pixel-based extraction of the pest; therefore, enabling proper classification and identification.

Seyit Kerimkhulle et al 2021 [10], introduces the significant issue of weed spread behavior and its impact on crop productivity in agricultural fields. This research, utilizing advanced remote sensing techniques, mainly multispectral satellite imagery, and LSTM neural networks, will focus on developing an effective crop-weed classification model to identify outbreaks that could be identified in the field at distances from 3-6 meters with accuracy levels as high as 94% to 96%. This work serves as a valuable reference in the field of weed management, showing information and approaches that could be utilized and developed further to aid in other research projects of the same type.

3. METHODOLOGY

This paradigm of deep learning emerged as a very powerful paradigm in machine learning that has the capability to revamp various domains since it can empower computers to learn complex patterns from vast amounts of data. Inside this realm of deep learning, several different models have been developed, each being adapted to address some specific kind of tasks and domains. These models find their expression within a wide range of architectures ranging from basic neural networks to complicated convolutional and recurrent structures. It is fully specialized for image processing and uses layers of convolution and pooling to learn features from the pixels. It is at its best with the like tasks of image classification and object detection. VGG16 and VGG19 can also be mentioned for their simple and easily effective architecture for image recognition. The use of transfer learning is really prevalent when there is scarce or limited labelled data, that means when adapting pre-trained models to new tasks. RCNNs outperform other object detection methods as it suggests what regions to check. Autoencoders are used for the main issues: dimensionality reduction and denoising; they are compressing data. Attention Neural Networks focus attention on relevant input elements; they have improved many selective attention tasks such as translation and image captioning.

This paper will give insight into several of the key deep learning models, including sequential models, CNNs, VGG16,

VGG19, transfer learning techniques, autoencoders, and attention neural networks. We are going to explore the features of the application domain of these models and show how important they are in current machine learning research and practical deployment.

3.1 Sequential model

The approach in this paper was to very thoroughly classify images based on the criterion of plant vs. weed. Input data, which represented the two-dimensional image data, were fed into the Flatten layer as preprocessed to be converted to a one-dimensional vectorized format in order to integrate well with later layers. The first layer in our neural network architecture contained 64 neurons that utilized the ReLU activation function. Input photos saw this layer as an excellent feature extractor, giving intricate patterns and representations. Then, adding the secondary layer with 32 neurons and ReLU activation function increased model's capability to identify fine hierarchies in the image data. Two neurons with the activation function such as Sigmoid formed the output layer that was deployed to differentiate the two classes: plant or weed. This is because the job of classification was two-classed. This configuration eased the construction of probabilistic predictions, assisting in the binary classification of the input pictures. Our model architecture is well-thought-out, efficiently converting two-dimensional picture input into a format that is appropriate for binary classification. The layer-by-layer integration of appropriate activation functions, which guarantee the extraction and exploitation of relevant data, is the basis for its capacity to discriminatively categorize weed and plant species.

The actual cross entropy loss function is employed in the training of the Sequential Model where it will give the difference between the actual ground-truth labels assigned to each input image from the predicted class probabilities. In this example, a variation of it was used referred to as Binary_Crossentropy since what we are trying to solve here is a classification problem and it is Binary. The weights or the parameters for which this model will make predictions are optimized iteratively using gradient descent and back-propagation. A loss function must be selected and optimized using some process in which the network's parameters are updated methodically. The optimization of this selection would thus ensure a proficient, non-stop evolution of the sequential model's efficacy and accuracy over time in a Classification-related ask.

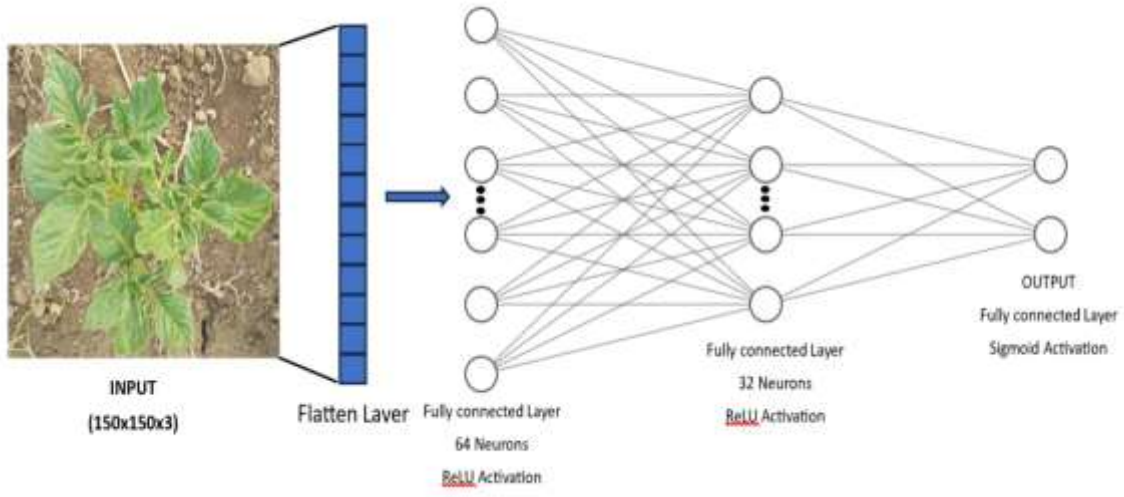

Fig. 1: Sequential Model Architecture.

3.2 CONVOLUTIONAL NEURAL-NETWORK (CNN)

One of the special neural networks used primarily in complex processing and analysis of visual input, particularly images, is called a convolutional neural network or CNN. Known for their excellent performance in various tasks like object identification, picture classification, and image recognition, a basic CNN architecture comprises distinct parts working cohesively to provide excellent performance. Convolutional Layers are the building blocks of a CNN. It applies the filtering, or kernels, convolutional operations very carefully to localized regions of the input image, preserving important spatial correlations and gathering local patterns and characteristics. Activation functions like ReLU, for example, also introduce some non-linearity after convolution. In other words, it allows a model to discover complex relationships and patterns in the data. The pooling layer is yet another crucial part of a convolutional neural network. The input volume's spatial dimensions are compressed using average or max-pooling algorithms, which reduces the likelihood of overfitting. The output from the pooling layers is flattened into a one-dimensional vector prior to being connected to Fully-Connected (Dense) Layers. In this way, it can be transferred easily from spatial hierarchies to a typical layout of a neural network. For providing the precise predictions to the Full Connected Layers, consolidation of the high-level characteristics achieved by earlier layers is crucial

to the model. Dropout layers Dropout layers can be deliberately introduced in order to control overfitting by making a subset of neurons inactive at random during the training process. It promotes reliance on alternative routes and prevents the model from becoming fixated on specific attributes. The final Output Layer combines the predictions of the model, applying softmax for multi-class situations and sigmoid function for the binary classification, depending on the nature of the job. Such adaptability and hyperparameter flexibility, in addition to the indisputable fact that fine-tuning is normally required for optimal working, illustrate why CNN structures are pertinent for applications as varied as classification to segmentation and, even more encompassing, visual data processing.

In the context of image classification, a well-engineered CNN was used with the aim of distinguishing between different species of plants and weeds. Our designed sequential CNN model began with a Conv2D layer having 64 filters of size (3, 3) is passed through Rectified Linear Unit (ReLU) function. When applied to images of size (150, 150, 3) this layer turned out to be an extremely strong feature extractor that could capture very complex patterns and representations. Then a MaxPooling2D layer with a window of (2, 2), and following it, BatchNormalization improved further the ability of the network to identify important characteristics and highlight them. Furthermore, this was complemented by adding another second Conv2D layer of 32 filters along the ReLU activation function followed by a second MaxPooling2D layer to obtain downsampling. Yet another additional Conv2D layer integrated in the CNN architecture flow made the features learnt even deeper as well as subtle. There were convolutional and pooling layers whose output was successfully streamed into a one-dimensional vector from the next layer referred to as Flatten without any complications in turning to the fully connected layers. The first Dense Layer, comprising 32 neurons, would get activated by ReLU which will help the network to recognize complex hierarchical structures in the image dataset. The last Dense layer, with the discriminative purpose of distinguishing between the plant and weed classes, was conceived for the binary nature of the classification problem. It is composed by two neurons whose activity is sigmoid. Thus it could generate probabilistic predictions, directly aiding in creating categories from input photos**.**

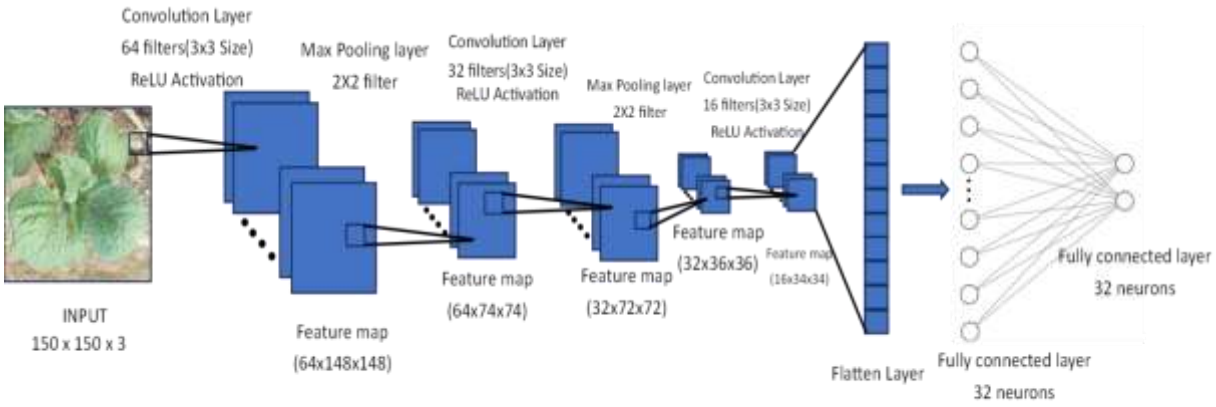

Fig. 2: Convolutional Neural-Network Model Architecture.

4. TRANSFORMERS

The ViT should be viewed as a revolution in the computer vision task space, as it challenged CNN's dominance by adopting a new architecture based on the success of transformers in NLP. It is different from CNNs, where local receptive fields and hierarchical feature extraction are considered by the convolutional layers, while ViT assumes that image understanding is based on global self-attention mechanisms. This way, ViT makes the model capture local as well as long-range dependencies in the image, thereby achieving better feature extraction and representation learning. The core of ViT is the self-attention mechanism ; it enables the model to give the relative importance of various patches within images. These will enable ViT to pay attention to the relevant parts of an image and suppress noise and other irrelevant information, which improves generalization and robustness. More importantly, using transformer blocks gives ViT opportunities to enhance the contextual information and semantic relationships between patches thus making the model viable to learn the details and structures in the images. For instance, Vision Transformers can be scaled and adjusted to a variety of image dimensions and ratio. Unlike CNNs at times require complex architectural transformations to facilitate a variety of inputs, ViT can take images of arbitrary size by slicing them into fixed-size patches and then applying a feed-forward network composed of several transformer layers. This intrinsic scalability makes ViT a very good candidate for a variety of vision tasks like segmentation, object detection, and picture classification, and even generative modeling. As such, the Vision Transformer has caused huge interest and adoption among the computer vision community as well.

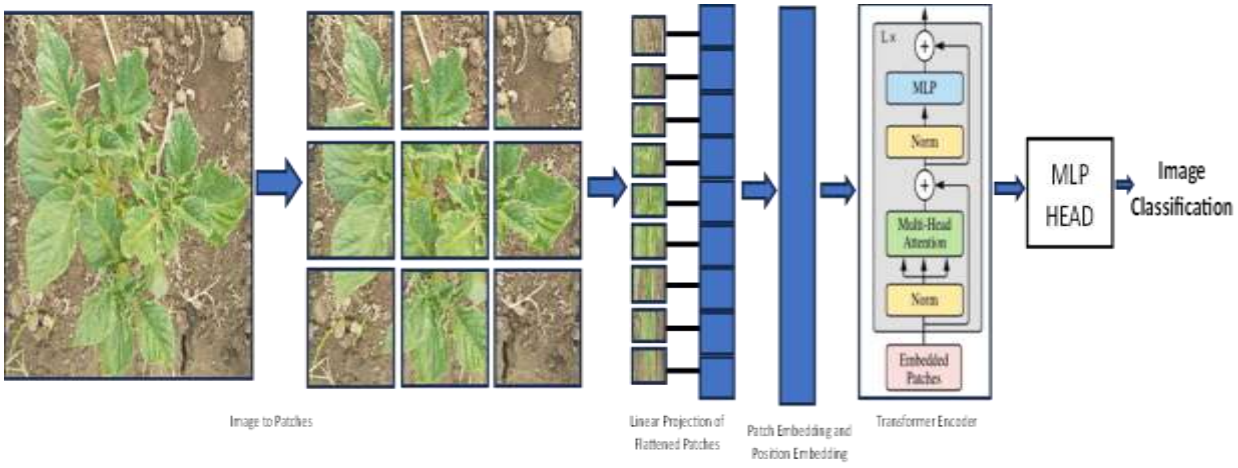

Fig. 3: Vision Transformer Model Architecture.

Determine the window type.

Vision Transformer Algorithm:

Step 1- Input Image Encoding: The input image is broken into multiple patches. Then, The patch is implanted into a lower-dimensional region in a linear fashion. The transformation serves to transform the patch into a fixed-size vector, which is the right way for transformer layers processing shown in equation 1

X=PatchEmbed(I) ---------equation(1)

Step 2 - Positional Encoding Since transformer architecture inherently does not have positional information such as images, the patch embeddings add positional encodings for encoding spatial relationships. As this is necessary for capturing spatial dependencies within an image Processing shown in equation 2

Xpos=X+PosEmbed(X)X_{pos} ----------equation(2)

Step 3- Multiple Transformer encoder layers are subsequently traversed by the patch embeddings using positional encodings. There are the two primary sub-layers that make up each encoder layer.

3.1 Multi-Head Self-Attention: that allows the model to handle several input components concurrently, the input is split up into many heads. The associations between patches are used by each head to calculate attention weights. Equation 3 displays the processing.

MultiHead(Q,K,V)=Concat($head_1$,...,$head_h$)W^O------equation(3)

3.2. Position-wise Feed-Forward Networks: Following self-attentionEach position is subjected to a separate position-wise feed-forward network in order to capture interactions between different dimensions of the embeddings Processing shown in equation4.

FFN(X)=ReLU(XW$_1$+b_1)W_2+b_2-------equation(4)

Step 4- Layer Normalization and Residual Connection: The model can learn more efficiently by removing the vanishing gradient issue after each sub-layer (feed-forward network and self-attention) is followed by layer normalization and a residual connection. The processing displayed in equation 5

LayerNorm(X+SubLayer(X))LayerNorm(X+SubLayer(X)) -----equation(5)

Step 5- Output:After a number of encoder layers, output embeddings are obtained. These are a collection of vectors that can be applied to downstream tasks including segmentation, object detection, and classification.

5. RESULT AND DISCUSSION

The implemented Deep Artificial Neural Network (ANN) model, using the Sequential architecture, was trained with image data resized to 28x28 pixels and augmented through various transformations such as shearing, zooming, and flipping. The network, comprising two dense layers (256 and 128 units), demonstrated effective learning during the training phase. The loss function as binary-cross entropy in stochastic gradient descent (SGD) based optimization allowed the model to efficiently minimize the error over epochs. The training and validation losses showed a steady decline, indicating reduced overfitting and a balanced generalization of the model. The accuracy metrics, as observed from the history plots, highlight

that the model successfully captured relevant features from the dataset, achieving competitive accuracy across both training and validation sets. These results underscore the capability of simple dense-layer networks performance on tasks such as image-based classification when combined with effective data augmentation techniques.

The Deep ANN model training provided meaningful outcomes, as seen in the accuracy and loss measures depicted in Figures 4 and 5. The accuracy of the model increased from a relative 54.76% in the initial epoch to 67.56% in the last epoch, proof of a steady learning pattern. This is graphically depicted in Figure 4, whereby the training accuracy consistently rises through the epochs, reflecting successful fitting to the training data. The loss during training trended downwards from 0.70 to roughly 0.64, a trend that reflects the accuracy enhancements and is evident in Figure 5. This demonstrates improved performance since the model tweaked its parameters over the course of training.

Validation metrics also present important information on model performance. While validation accuracy varied, the highest value at 78.67% occurred at the 16th epoch. This implies that while the model generalized more effectively at this stage, there were instances of overfitting, especially for later epochs. For example, at epoch 11, the validation accuracy is 74.67%, whereas training accuracy is much greater at 59.52%, pointing to a divergence that manifests the necessity for early stopping or regularization to promote generalization.

In addition, the loss for validation also had a fluctuating pattern, going down to 62.62 at epoch 19, indicating the enhanced capacity of the model in predicting the unseen data. The interaction between the training and validation metrics, as explained and illustrated in the plots, also highlights the need to maintain balance to ensure that the model learns to recognize generalizable characteristics. In general, these findings, as shown in Figures 4 and 5, illustrate the efficacy of the Deep ANN architecture, highlighting its ability to learn intricate patterns in image data while also pointing to areas for future optimization and improvement.

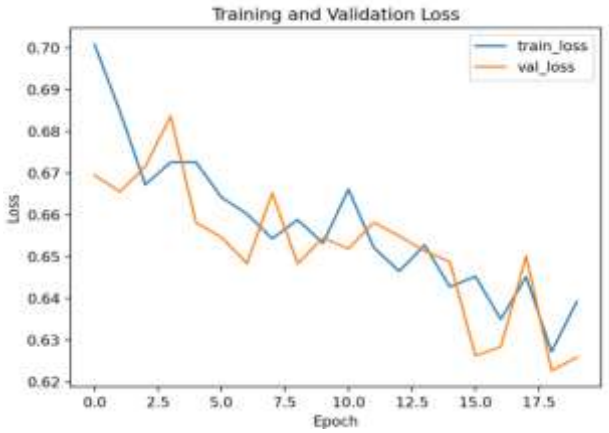

Fig 4: Training Accuracy and loss of Sequential Model

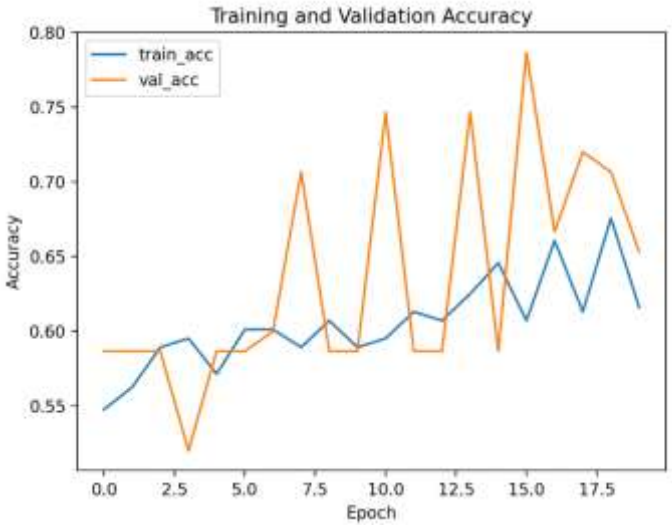

Fig 5: Validation Accuracy and loss of Sequential Model

Convolutional Neural Network (CNN) was designed for identifying the carrot crops and weeds based on 150x150-pixel resized images. The model has several convolutional layers with 64 and 32 kernals, using ReLU activation to discover intricate patterns. MaxPooling2D is applied after each convolutional layer to decrease spatial size and assist in feature extraction. Batch Normalization is added to improve training stability.

The architecture finishes with two densely connected Dense layers, as it is binary classification problem the last layer uses sigmoid activation function. The model was trained with data augmentation strategies, including shearing and flipping, to enhance generalization. With binary cross-entropy and the Adam optimizer, the training history indicates successful learning over epochs. In total, these findings prove the robustness of the architecture and its appropriateness for separating carrot plants from weeds and enabling more effective agricultural practice.

The presision of the CNN model in the classification of carrot plants and weeds was tested for 20 epochs, as presented in figure of the paper. The model showed considerable improvement in training and validation accuracy across the epochs. Beginning with a training accuracy of around 45.5% in the initial epoch, the model reached an impressive 100% accuracy by the fifteenth epoch. Similarly, the training loss went down from 0.979 in the initial epoch to 0.006 by the last epoch, reflecting successful learning and convergence.

Accuracy in validation first started at about 41.3%, rising slowly to about 97.3% by the nineteenth epoch, while the loss in validation reduced from 0.698 to 0.206. All these trends are graphically shown in Figure 6 and 7, depicting the reduction in validation and training loss, and rise in accuracy in epochs. The plots clearly indicate the model's generalization well to unseen data, with validation scores constantly improving together with training scores. Generally, the results reflect the stability of the architecture, highlighting its ability to precisely identify carrot plants from weeds. This points to the model being ideal for agricultural use, allowing improved control and decision-making in crop farming.

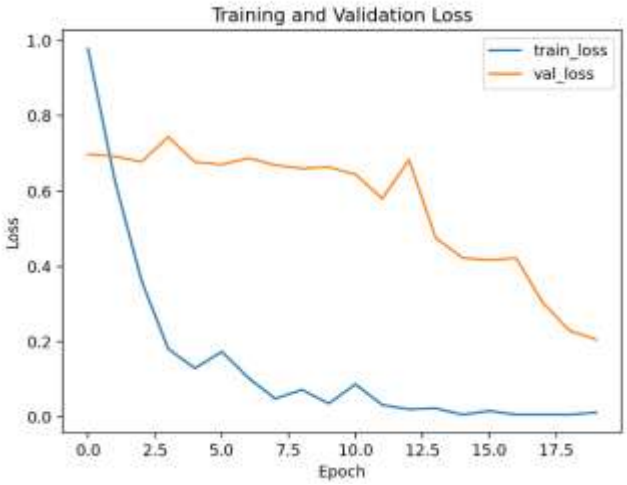

Fig 6: Training Accuracy and loss of CNN Model

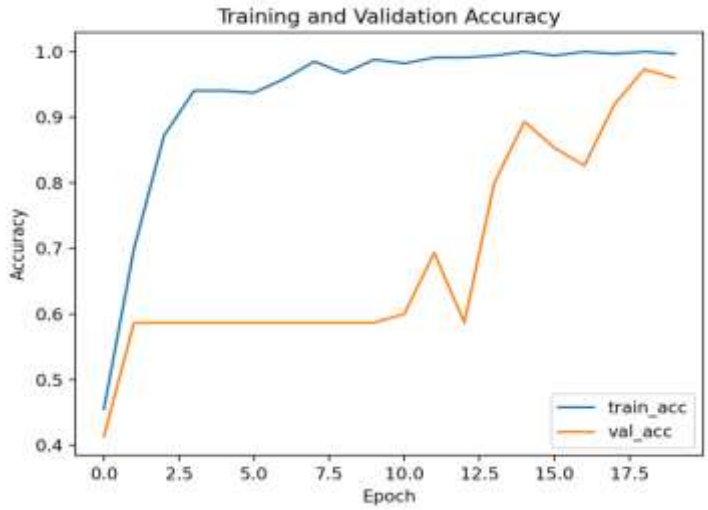

Fig 7: Validation Accuracy and loss of CNN Model

a Vision Transformer (ViT) architecture for plant and weed classification, leveraging its capacity to develop good models of relationships between image patches, which is useful for inferring complex visual information. We used a pre-trained ViT model (ViT-B/16) from the torchvision package, loaded with weights trained on a large dataset. This transfer learning method enables the model to utilize previously acquired features, improving performance on our particular classification task. The architecture of the model involves a frozen base feature extractor where only the classifier head is trainable, enabling efficient training while maintaining the strong representations learned from the pre-trained model. We changed the last layer for classifying two classes—daisy and dandelion—accurately indicating our requirement of separating carrot plants from weeds. The model was trained with cross-entropy loss and optimised using the Adam optimizer, showing successful learning during the training phase.

Throughout 10 epochs, the model recorded a great improvement in accuracy, with the final training accuracy being 100% and validation accuracy being 100%.

Simultaneously, the training loss also reduced substantially, from 0.4325 in the initial epoch to 0.0264 in the last epoch, and the validation loss also reduced from 0.2492 to 0.0186 . All these trends are graphically shown in Figure 8 and 9. These outcomes reflect the model's capability to generalize well to unseen validation data. The observed trends in the training loss and validation loss, along with the corresponding accuracy improvements, are also depicted in the loss and accuracy graphs presented in the paper. The Vision Transformer architecture enabled efficient feature extraction from image patches, which is well-adapted to high-resolution images such as those in our database. The model's capacity to handle complex spatial interactions among image patches resulted in its impressive ability to identify between carrot plants and weeds, further validating the potential of transformer-based models in agriculture.

These results show that the ViT architecture, owing to its powerful learning ability, presents a good solution for precision agriculture, with the potential for accurate weed detection and classification. This, subsequently, can lead to optimized management of weeds and less herbicide use, pushing forward sustainable farming methods.

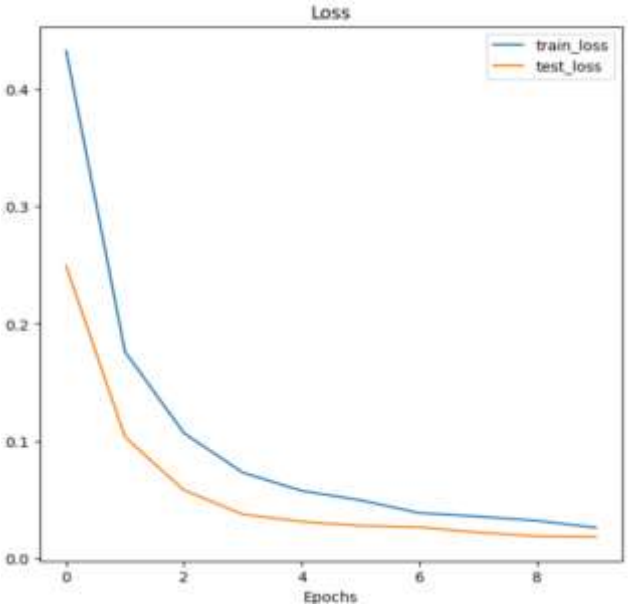

Fig 8: Training Accuracy and loss of ViT Model

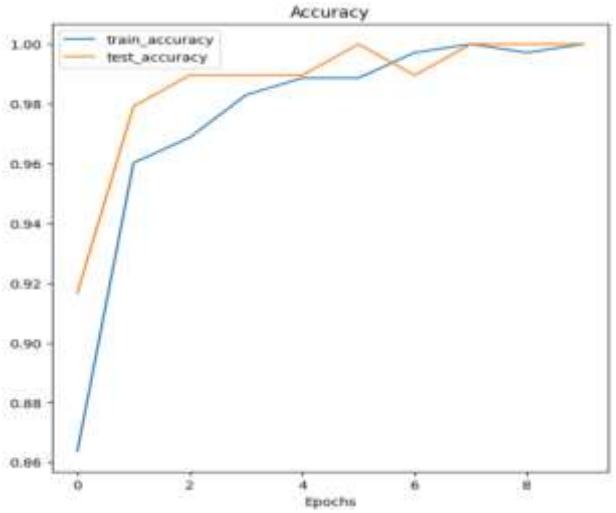

Fig 9: Validation Accuracy and loss of ViT Model

REFERENCES

[1]] J. Sangeetha a, Priya Govindarajan b, "Prediction of agricultural waste compost maturity using fast regions with convolutional neural network(R-CNN)", Materialstoday Proceedings, DOI: https://doi.org/10.1016/j.matpr.2023.01.112, Published: 2023.

[2] [2] Nanyang Zhu, Xu Liu, Ziqian Liu, Kai Hu, Yingkuan Wang, Jinglu Tan, Min Huang, Qibing Zhu, Xunsheng Ji, Yongnian Jiang, Ya Guo, "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities",International Journal of Agricultural and Biological Engineering, Vol 11, No 4 (2018), DOI: 10.25165/j.ijabe.20181104.4475, Published on: 2018.

[3] [3] Muhammad Hammad Saleem 1,Johan Potgieter 2 andKhalid Mahmood Arif 1, Plant Disease Detection and Classification by Deep Learning", Plants 2019, 8(11), 468; https://doi.org/10.3390/plants8110468, Published: 31 October 2019

[4] S. Kavitha, T. Satish Kumar, E. Naresh, Vijay H. Kalmani, Kalyan Devappa Bamane & Piyush Kumar Pareek, "Medicinal Plant Identification in Real-Time Using Deep Learning Model", SN COMPUT. SCI, DOI: https://doi.org/10.1007/s42979-023-02398-5, Volume 5, article number 73, Published: 07 December 2023.

[5] Jun Liu & Xuewei Wang, "Plant diseases and pests detection based on deep learning: a review", Plant Method 17,Article number:22, DOI: https://doi.org/10.1186/s13007-021-00722-9, Published: 2021.

[6] Alvaro Fuentes, Sook Yoon & Dong Sun Park, "Deep Learning-Based Techniques for Plant Diseases Recognition in Real-Field Scenarios", Springer , Volume 12002, ISBN : 978-3-030-40604-2, Published on: 06 February 2020

[7] Sunil G C a, Yu Zhang a, Cengiz Koparan a, Mohammed Raju Ahmed a, Kirk Howatt b, Xin Sun a, "Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions", Journal of Agriculture and Food Research, Volume 9, DOI:10.1016/j.jafr.2022.100325, Published on: September 2022

[8] Senthil Kumar Swami Durai a, Mary Divya Shamili b, "Smart farming using Machine Learning and Deep Learning techniques",Decision Analytics Journal,Volume 3, DOI: https://doi.org/10.1016/j.dajour.2022.100041, Published on: June 2022.

[9] Niranjan C Kundur; P B Mallikarjuna, "Pest Detection and Recognition: An approach using Deep Learning Technique", 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), pp. 1-6, DOI: 10.1109/CCIP57447.2022.10058692, Published: 2022

[10] Seyit Kerimkhulle; Zhandos Kerimkulov; Dias Bakhtiyarov; Nazerke Turtayeva; Jong Kim, "In-Field Crop-Weed Classification Using Remote Sensing and Neural Network", pp. 1-6, DOI: 10.1109/SIST50301.2021.9465970, Published in 2021.

/