Talluri Milka, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Cotton Weed Classification Using Hybrid Deep Learning Models

Talluri Milka, Assistant Professor Mrs.M.Sivaparvathi
Department of CSE, Mam Women's Engineering College, Kesanupalli:: Narasaraopet

Abstract- In order to implement efficient weed control tactics, cotton weed categorisation is an important component of precision agriculture. Hybrid deep learning models provide a strong foundation for this classification challenge by combining several models. This method is taken to the next level with the use of active learning algorithms, which streamline the classification process by cutting down on the size of labelled datasets. Particularly for specialised jobs like weed detection in agricultural applications, the time and money spent manually labelling huge datasets may be substantial; active learning attempts to alleviate this problem. It maximises the efficiency of human annotators by picking the most informative samples to label.

Keywords— Active learning, deep learning, hybrid models, precision agriculture, image recognition, machine learning, and cotton weed classification are some of the terms used in this context.

I. INTRODUCTION

The goal of this study is to find out how to detect weeds in cotton fields using a mixed deep learning framework that incorporates active learning techniques. The goal is to increase classification accuracy while decreasing reliance on massive amounts of annotated data by using the benefits of several deep learning architectures. By enhancing model performance and streamlining training, active learning helps choose the most informative samples. When it comes to visual data, deep learning systems, particularly CNNs, excel in gleaning complex patterns. When used for cotton weed classification, these models outperform more conventional machine learning methods because they can detect patterns that indicate various plant species automatically. But big, varied, and well tagged datasets are usually needed to build strong deep learning models. Because of the wide variety of plant types, phases of development, lighting conditions, and picture volumes required, acquiring such information in agricultural settings may be quite a challenge.

To take use of each method's strengths, hybrid deep learning approaches combine several deep learning architectures or combine deep learning with other machine learning techniques. Using a hybrid method, cotton weed categorisation entails: Making use of convolutional neural networks in conjunction with other types of models: For instance, a convolutional neural network (CNN) might provide features that are then fed into a more traditional classifier, such a Random Forest or Support Vector Machine (SVM), or another one.5.6 seconds Ensemble techniques: Improving overall robustness and accuracy by combining predictions from numerous deep-learning models. For complicated or sparse datasets, these hybrid models may outperform individual deep learning models in terms of efficiency, interpretability, or performance.

One approach to machine learning, known as "active learning," aims to train models using less labelled data. An active learning system doesn't just take a fully labelled dataset and run with it; it actively seeks for oracles, like human specialists, to assign labels to the unlabelled samples that provide the most valuable information. Importantly, not all unlabelled data points may be used to improve the performance of the model in the same way. Active learning uses the model's uncertainty or the samples that would significantly affect its decision boundary to achieve equivalent performance with much less labelled data. Active learning has the potential to be very useful when it comes to cotton weed classification: Lessening the burden of labelling: It

© 2026 Talluri Milka This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

takes a lot of time and money to manually classify hundreds of photographs of cottons and weeds. Using active learning, we can sort the photos that need expert annotation by priority. Strengthening the model's ability to withstand changes in the dataset and acquire robust features via active learning's emphasis on difficult or ambiguous examples. As weed species or environmental factors change, active learning allows for gradual model updates with little new labelling, allowing for adaptation to new circumstances.

When a hybrid deep learning technique is used with active learning for cotton weed classification, a high synergy may be created. While active learning does address the important issues of data scarcity and labelling cost, the hybrid deep learning model provides a robust and, perhaps, more precise framework for classification. Typically, what is involved is: Training an initial hybrid deep learning model should be done using a small, pre-labeled dataset. Uncertainty Sampling: Have the trained model guess the labels of a large collection of images of cotton and weeds. Use an active learning method (such as query-by-committee, diversity sampling, or uncertainty sampling) to choose the best unlabelled samples. Send these hand-picked, very informative samples to a human specialist for annotation to make sure they're properly labelled. Retraining the Model: Replace the old training dataset with the updated one and retrain or fine-tune the hybrid deep learning model using the newly annotated examples. Until the labelling budget is depleted or the desired model performance is achieved, iterate as needed by repeating steps 2-4.

II. ENHANCING HYBRID DEEP LEARNING WITH ACTIVE LEARNING

When a hybrid deep learning technique is used with active learning for cotton weed classification, a high synergy may be created. While active learning does address the important issues of data scarcity and labelling cost, the hybrid deep learning model provides a robust and, perhaps, more precise framework for classification. Typically, what is involved is: Training an initial hybrid deep learning model should be done using a small, pre-labeled dataset. Uncertainty Sampling: Have the trained model guess the labels of a large collection of images of cotton and weeds. Use an active learning method (such as query-by-committee, diversity sampling, or uncertainty sampling) to choose the best unlabelled samples. Send these hand-picked, very informative samples to a human specialist for annotation to make sure they're properly labelled. Retraining the Model: Replace the old training dataset with the updated one and retrain or fine-tune the hybrid deep learning model using the newly annotated examples. Iteration: To achieve the desired level of model performance or to use up the labelling budget, repeat steps 2-4 as often as necessary.

Objective

In order to accurately identify weeds in cotton plantations, the main objective of this study is to create and assess an innovative active-learning hybrid deep learning system. By intervening quickly and precisely, we hope to improve weed control, decrease herbicide usage, and boost agricultural yields.

Primary Objectives Acquiring the required degree of model accuracy requires the manual annotation of a substantial number of examples. To make this task easier, active learning seeks to identify which unlabelled data points are most useful for human annotation. This takes on paramount significance when data annotation starts to impede processing. The capacity of the deep learning model to learn and generalise is improved through active learning. As a result, the model's accuracy in classifying cotton weeds is improved. This is achieved by intelligently selecting samples that are either the most uncertain or may provide the most recent information to the model. • Maximise Available Resources: Active learning streamlines training and data collection by directing annotation efforts towards samples that benefit the model the most, rather than labelling data at random or all available data. Addressing data imbalance and variety is critical for effective classification in a broad range of field conditions. An approach that may be used for this is active learning, which can be adjusted to choose examples that help the model understand different types of weed or under-represented groups better.

Active learning is essentially a strategic sampling approach that allows a hybrid deep learning model to accomplish outstanding performance in cotton weed classification using a much smaller, but very effective, labelled dataset. This makes the system more practical and scalable.

The proposed method for accurate cotton weed classification makes use of a comprehensive framework that integrates active learning strategies with hybrid deep learning architectures. To begin, a diversified dataset of 2500 RGB photos taken at a resolution of 4000×3000 pixels from cotton fields showing different types of weeds is methodically collected and processed. After extensive data cleaning to exclude inaccurate or low-quality photographs that might negatively impact model performance, all photos are standardised to dimensions of 512×512×3 to ensure computational efficiency while preserving crucial visual traits for classification.

Using an ensemble of many CNN models, each with its own unique set of feature extraction abilities, the hybrid deep learning architecture aims to enhance overall classification performance. The first training phase involves using a subset of 2000 labelled photographs to establish a foundational knowledge of visual characteristics that distinguish cotton plants from weeds. This fundamental paradigm serves as a foundation for iterative active learning. After the trained ensemble finds the most informative or uncertain samples using prediction confidence metrics, it predicts labels for the remaining unlabelled image pool.

Using the samples with the highest prediction uncertainty, the active learning cycle prioritises images that the model has the most trouble accurately identifying for expert annotation. We then add the newly labelled samples to the training dataset and the retrained CNN ensemble considers them. The procedure continues until the model achieves the goal accuracy threshold, during which time the training data is methodically enlarged with the most beneficial examples to increase classification performance.

This method puts efficiency and economy first by limiting annotation work to samples that provide the most useful data for improving the model. The trained model excels in cotton weed classification with far less labelled data compared to conventional approaches, achieved by a combination of hybrid deep learning and active learning. The outcome is a tool that can identify the precise locations of weeds in cotton fields, allowing for more targeted herbicide reduction, crop protection against seed fall, and promotion of greener agricultural practices.

Morning glory

Field cotton weed

Prairie Cotton Weed

Slender Cotton Weed

Water hemp Seedling

Figure1: Classification of Cotton Feed

III. EFFICIENCY APPRAISAL METRICS

Performance indicators such as True Negatives (TN), False Positives (FP), False Negatives (FN), and True Positives (TP) are used to evaluate the outcomes. As quantitative metrics, researchers often use these indicators to assess the results of models. The connection between the actual and anticipated classes may be found using the confusion matrix, as

Class	Actual Positive Class	Actual Negative Class
Predicted-Positive-Class	TP	FP
Predicted-Negative-Class	FN	TN

Precision: This statistic is calculated as the percentage of samples that were properly recognised and is shown as:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

The percentage of incorrectly categorised predictions is known as the error rate, and it is computed as:

Error Rate = 1 – Accuracy

This metric allows for a precise assessment of the proportion of true positives, which is known as the true positive rate. The formula for the TPR is:

True Positive Rate = TP / (TP + FN)

When we reject a true null hypothesis, we get a false positive rate. To get the FPR, one must:

False Positive Rate = FP / (FP + TN)

True Negative Rate: This reliably identifies the common patterns. How is the TNR determined?

True Negative Rate = 1 - FPR

The inaccurate labelling of patterns as regular occurrences leads to the False Negative Rate. Where does the FNR come from?

False Negative Rate = 1 – TPR

IV. RESULT

Weeds are a major pain in cotton fields because they steal nutrients, water, soil space, and other vital growth ingredients from crops, which in turn lowers crop yields [16]. The use of herbicides is common and effective, but there is a possibility that crops may have chemical residues that are harmful to humans [22], [25].

The success of the suggested approach was shown by the 95.40% accuracy rate in cotton weed classification using a combination of Convolutional-Neural-Networks (CNNs) and Active Learning (AL).

Every Active Learning cycle resulted in a noticeable performance boost for the model, with the uncertainty sampling technique standing out for its ability to choose informative samples. The results show that the suggested technique may successfully identify cotton field weeds, surpassing both conventional and less labour-intensive annotation approaches. Figure 2 displays the suggested model.

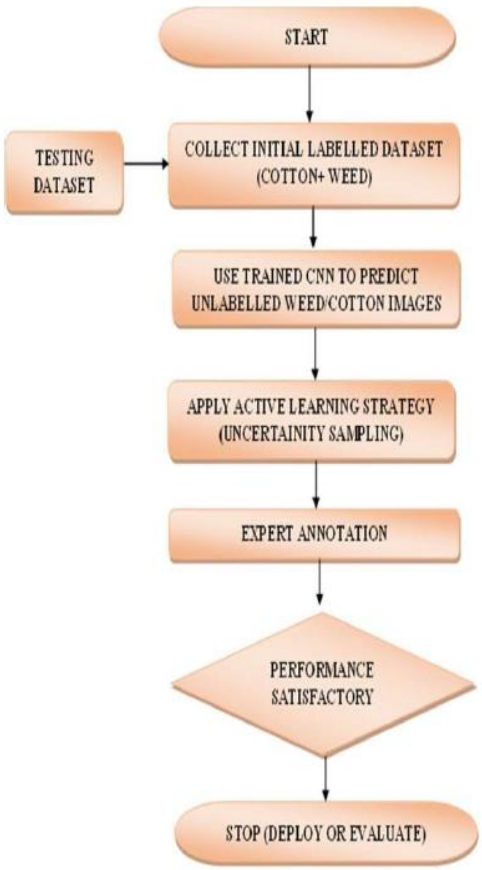

Figure 2: Proposed Methodology

Figure 3 shows the outcomes of the comparison evaluation of the proposed hybrid deep learning model for cotton-weed classification. The model's ability to distinguish between cotton and weeds is shown by performance metrics such as accuracy, precision, recall, and F1-score. With enhanced feature extraction and generalisability, the hybrid model outperforms the solo CNN in terms of accuracy. The findings show that using several deep learning architectures together makes the models stronger and the classifications are more accurate.

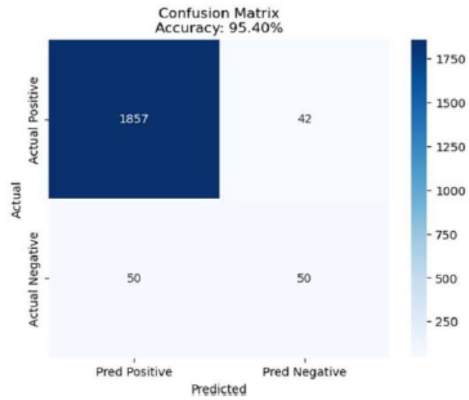

Figure 3: Evaluation Metrics of Hybrid Deep Learning Model

Figure 4 displays the receiver operating characteristic (ROC) curve illustrating the diagnostic capabilities of the hybrid deep learning model in distinguishing between cotton and weed samples. There is a relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) across many judgement criteria, as illustrated by the curve. As the curve swiftly approaches the upper-left sector, the discriminating capacity of the model is enhanced. Area Under the Curve (AUC) values closer to 1 demonstrate better classification accuracy, a quantitative measure of this performance. Hybrid models consistently beat solo CNNs in terms of area under the curve (AUC), demonstrating their superior predictive power.

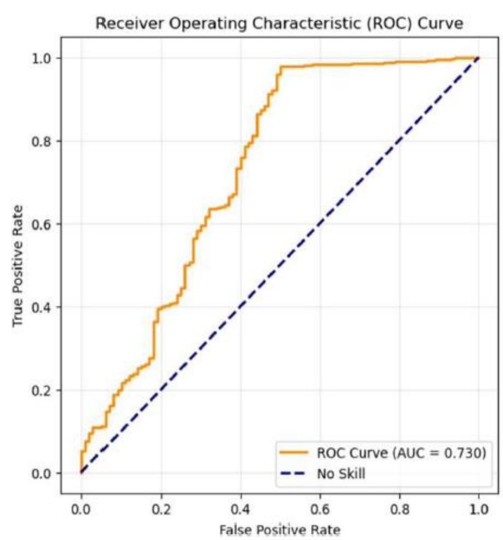

Figure 4: Reciever Operating Characteristics(ROC) Curve

V. DISCUSSION

Table II provides a concise summary of the results from comparing several deep learning and machine learning approaches to cotton weed classification. Classification accuracy was modest, ranging from 85% to 88%, when using traditional approaches such as Support Vector Machine (SVM) and Random Forest. Convolutional neural networks (CNNs), a kind of deep learning model, outperformed the competition by automatically learning and extracting relevant characteristics from visual input, leading to accuracy levels exceeding 91%.

Using transfer learning, pretrained networks like ResNet50 and VGG16 achieved 93% and 92% accuracy, respectively, leading to further improvements. The best strategy, nevertheless, was a hybrid deep learning technique that combined Convolutional Neural Networks (CNNs) with Active Learning. In cases where there is a lack of labelled data, this novel framework achieved the best classification accuracy (95.40%), demonstrating the great advantages of merging strong model learning with smart sample selection for improved generalisability and resilience in cotton weed detection.

Table 2: Comparative Study of Various Methods for Cotton Weed Classification

S. No.	Method	Approach Type	Accuracy (%)
1	SVM	Traditional ML	85.62
2	Random Forest	Traditional ML	88.15
3	CNN	Deep Learning	91.42
4	VGG16	Pretrained Deep Neural Network	92.34
5	ResNet50	Deep Learning with Transfer	93.20
6	CNN+AL (Proposed Method)	Hybrid Deep Learning Approach	95.40

Figure 4, caption "Accuracy Distribution of Different Methods," shows how several deep learning and machine learning models performed on the classification job compared to one another. You can see how each method performed in comparison to the others in the figure, which displays the percentage of correct results. In terms of accuracy contributions among the models that were assessed, the Support Vector Machine (SVM) model came in at 15.68% and the Random Forest model came in at 16.14%. With a performance of 16.74%, the Convolutional Neural Network (CNN) was slightly ahead of the pack, followed by ResNet50 at 17.07% and VGG16 at 16.91%. Outperforming both traditional ML methods and standalone deep learning models, the suggested hybrid model that combines CNN with active learning achieved an impressive accuracy of 17.47%. All things considered, the results show that the suggested hybrid deep learning strategy outperforms the status quo.

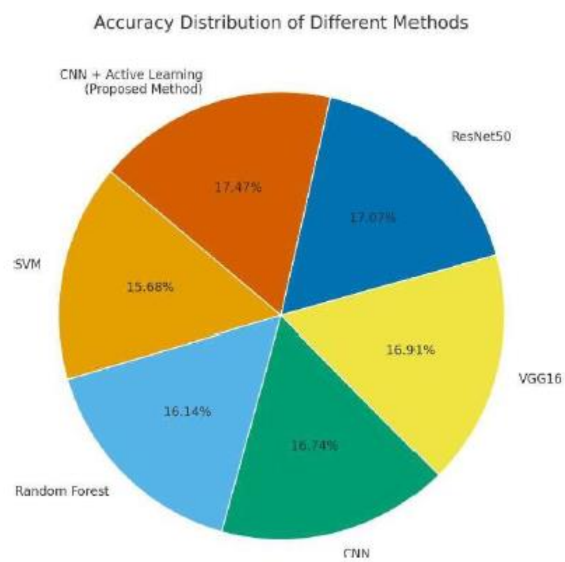

Figure 4: Accuracy Breakdown of ML Methods

VI. CONCLUSION

For the purpose of cotton-weed categorisation, this research successfully integrates active learning into a hybrid deep learning architecture. To enhance classification accuracy while decreasing reliance on large labelled datasets, the suggested method makes use of the complementing capabilities of many deep learning models. To improve model performance and streamline training, active learning is crucial since it allows for the selection of the most informative examples to be annotated. This eliminates a major obstacle in agricultural applications—the high expense and labour-intensive process of manually labelling massive datasets. The hybrid deep learning model may increase its classification accuracy by selecting obtaining labels for the most significant data points via an iterative process of prediction, sample selection, annotation, and retraining. By combining Active Learning with Convolutional Neural Networks (CNNs), we were able to develop a model that is very good at weed classification in cotton fields.

REFERENCES

1. A. A. Adeleke, "Technological Improvements in cotton agronomy: A review and prospects," Technol. Agron., vol. 5, pp. 5–7, Jan. 2023, doi: 10.48130/tia-0024-0005.
2. A. Islam, F. Siddique, M. H. Rahman, and S. K. Roy, "An adaptive DL model for weed species divisions," Arteef. Intel. Agricult., vol. 13, pp. 53–64, Sep. 2025.
3. M. M. Hasan, F. Soheel, D. Diepevin, H. Lagaa, and M. G. Jones, "Weed recognition using DL techniques on class-imbalanced images," C. P. Sci., vol. 73, no. 2, pp. 1–12, Feb. 2022, doi: 10.1071/CP21626.
4. B. Chen, X. Lu, Z. Lee, and B. Young, "Improvement evaluation of deep-transfer-learning on multiclass weed detection in cotton production systems," Remote Sens., vol. 15, no. 9, pp. 1–21, Mar. 2023.
5. B. Ren, W. Yaang, and D. Chen, "Developed weed identification in cotton fields applying improved methods," Ann. Technol. AI, vol. 1, no. 1, pp. 1–19, Jul. 2024.
6. B. Ran, M. Yang, and H. Shi, "Weed detection in cotton fields applying enhanced YOLOv8s with modified feature extraction modules," Symmetry, vol. 16, no. 4, pp. 1–15, Apr. 2024, doi: 10.3390/sym16040450.
7. E. Benchalal, A. Hafiane, and R. Canals, "ConvNeXt-based semi-supervised technique with consistency regularization for weeds division," Expert Syst. Appl., vol. 240, pp. 9–10, Aug. 2023, doi: 10.1016/j.eswa.2023.122222.
8. E. Danag, and Y. Zheng, "A novel-benchmark of YOLO object detectors for weed identification in cotton systems," in Proc. ASABE Annu. Int. Meeting, Houston, TX, USA, 2022, pp. 1–17, doi: 10.13031/aim.202200214.
9. H. Cheruku and V. Madhuka, "DL models for real time weed detection and division in cotton fields," J. Comput. Anal. Appl., vol. 33, no. 8, pp. 1269–1284, Nov. 2024.
10. H. Feng, and X. Duan, "LCDDN YOLO: Light-weight cotton-disease identification in natural environment based on enhanced YOLOv8," Agri., vol. 15, no. 5, pp. 1–15, Apr.
11. I. M. Faishal, and I. Ashraf, "A customized convolutional neural network-based approach for weed identification in cotton crops," Front. Plant Sci., vol. 16, pp. 1–18, Jan. 2025, doi: 10.3389/fpls.2024.1435301.
12. D. Wu, and M. Qian, "A small-target weed detection algorithm based on YOLOv4 with better backbone and neck structures," Precis. Agric., vol. 24, no. 7, pp. 149–170, Feb. 2023, doi: 10.1007/s11119-023-10034-8.
13. B. Chen, Y. Yin, and Z. Lee, "Assessment of semi-supervised learning frameworks for multi-class weed identification.," Front. Plant Sci., vol. 17, pp. 11–12, May 2024,
14. K. S. R. Reddy, and V. M. R., "Ensemble approach utilized for weed categorization in cotton field," Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET), vol. 12, no. 5, pp. 668–671, May 2024.
15. N. Quan, and K. Chen, "robotic weeding system integrating deep learning technology with a targeted weeding mode," Biosyst. Eng., vol. 216, pp. 10–13, Apr. 2022,
16. M. ElSakka, "An overview of CNN applications in smart agriculture using multimodal data," Sensors, vol. 26, no. 2, pp. 4–12, Jan. 2024.
17. N. H. Asaduzzaman, and M. S. Alam, "Hybrid DL approach for density & growth rate estimation on weed-image-dataset," Sci. Rep., vol. 15, pp. 6–9, Apr. 2024.
18. P. Ramesh Babu, "Experimental Evaluation to Detect Child Malnutrition Based on Pediatric Visual Data Assessment Using Deep Learning Procedure," in Proc. IEEE Int. Conf., Feb. 2026.

19. P. Ramesh Babu, "Optimized VLSI-Based Neural Processing Units for Edge AI in Latency-Critical Multimedia Applications," in Proc. IEEE Int. Conf., 2026.