ORIGINAL ARTICLE

Class-specific data augmentation for plant stress classification

Nasla Saleem[1] | Aditya Balu[1] | Talukder Zaki Jubery[1] | Arti Singh[2] |
Asheesh K. Singh[2] | Soumik Sarkar[1] | Baskar Ganapathysubramanian[1]

[1]Department of Mechanical Engineering, Iowa State University, Ames, Iowa, USA

[2]Department of Agronomy, Iowa State University, Ames, Iowa, USA

Correspondence
Baskar Ganapathysubramanian, Department of Mechanical Engineering, Iowa State University, Ames, IA, USA.
Email: baskarg@iastate.edu

Assigned to Associate Editor Prof. Jennifer Clarke.

Abstract

Data augmentation is a powerful tool for improving deep learning-based image classifiers for plant stress identification and classification. However, selecting an effective set of augmentations from a large pool of candidates remains a key challenge, particularly in imbalanced and confounding datasets. We propose an approach for automated class-specific data augmentation using a genetic algorithm. We demonstrate the utility of our approach on soybean [*Glycine max (L.) Merr*] stress classification where symptoms are observed on leaves; a particularly challenging problem due to confounding classes in the dataset. Our approach yields substantial performance, achieving a mean-per-class accuracy of 97.61% and an overall accuracy of 98% on the soybean leaf stress dataset. Our method significantly improves the accuracy of the most challenging classes, with notable enhancements from 83.01% to 88.89% and from 85.71% to 94.05%, respectively. A key observation we make in this study is that high-performing augmentation strategies can be identified in a computationally efficient manner. We fine-tune only the linear layer of the baseline model with different augmentations, thereby reducing the computational burden associated with training classifiers from scratch for each augmentation policy while achieving exceptional performance. This research represents an advancement in automated data augmentation strategies for plant stress classification, particularly in the context of confounding datasets. Our findings contribute to the growing body of research in tailored augmentation techniques and their potential impact on disease management strategies, crop yields, and global food security. The proposed approach holds the potential to enhance the accuracy and efficiency of deep learning-based tools for managing plant stresses in agriculture.

1 | INTRODUCTION

Accurate classification of plant stresses is of utmost importance for effective crop management and sustainable agricultural practices (Al-Hiary et al., 2011). Both biotic (diseases and insects) and abiotic plant stresses (drought, salinity, temperature extremes, and nutrient deficiencies) have detrimental effects on crop growth, yield, and quality (Mosa et al., 2017). By precisely identifying and classifying these stresses early, farmers can develop targeted strategies to mitigate their impact and optimize crop health (Nagasubramanian et al., 2018; Sankaran et al., 2010). Moreover, accurate stress classification plays a key role in selecting stress-tolerant crop varieties (Singh et al., 2021) and can make

a significant impact on improved genomic studies and high-throughput phenotyping (Singh et al., 2016, 2018; Zhang et al., 2017). Accurate stress classification can enhance cyber-agricultural systems, leading to improved crop resilience, reduced production losses, and sustainable agricultural practices (Gao et al., 2020; Gill et al., 2022; Gonzalez Guzman et al., 2022). In this paper, we explore the development of accurate classifiers for plant stress classification, aiming to improve downstream plant stress management activities involving stress identification and enable effective mitigation strategies.

Traditionally, plant stress identification and quantification heavily relied on the expertise of human scouts and domain experts (Singh et al., 2016). However, this manual approach is time-consuming, subjective, and limited in scalability, posing challenges in terms of efficiency and accuracy. The emergence of advanced technologies, such as drones (Feng et al., 2021; Guo et al., 2021; Herr et al., 2023; Xu et al., 2023), ground robots (Atefi et al., 2021; Gao et al., 2018), and sensors (Parmley et al., 2019; Pieruschka & Schurr, 2019) has brought high-throughput phenotyping and phenomics to the forefront (Araus & Cairns, 2014), transforming the measurement of multiple plant traits across various growth stages and facilitating rapid, precise, and accurate data collection. Machine learning (ML) and deep learning (DL) techniques have emerged as effective tools in automating plant stress classification processes (Ghosal et al., 2018; Singh et al., 2016). Despite promising outcomes in discerning various plant stresses, DL models encounter a significant challenge: the requirement for abundant labeled and diverse data (Kamilaris & Prenafeta-Boldú, 2018). To address this challenge, data augmentation (DA) has emerged as a valuable approach to enhancing model performance by augmenting the available data through various transformations (Krizhevsky et al., 2012; Shorten & Khoshgoftaar, 2019; Van Dyk & Meng, 2001). These transformations include rotation, flipping, scaling, cropping, and noise injection, effectively minimizing performance gaps between training and testing stages, reducing overfitting, and improving the generalization capability of DL models (Rebuffi et al., 2021; Shorten & Khoshgoftaar, 2019; Taylor & Nitschke, 2018a). Importantly, data augmentation allows for effectively expanding the training data without the need for laborious manual labeling or extensive data collection efforts, making DL models more accessible and efficient for plant stress classification tasks (Taylor & Nitschke, 2018a).

Despite the effectiveness of data augmentation in enhancing the performance of DL models, manually selecting appropriate augmentation techniques is time-consuming and challenging. To address this issue, researchers have turned to automated machine learning (AutoML) (He et al., 2021) techniques for automatically searching and selecting augmentation policies on datasets (Cubuk et al., 2018; Ho et al., 2019;

Core Ideas

- We propose an effective approach for automated selection of class-specific data augmentations for precise plant stress classification.
- Employing a genetic algorithm for efficient augmentation strategy selection in challenging datasets.
- Achieving significant performance gains with reduced computation via fine-tuning only the linear layer of the convolutional neural network model.

Lim et al., 2019; Zoph et al., 2020). These include methods like AutoAugment (Cubuk et al., 2018), Fast AutoAugment (Lim et al., 2019), and Faster AutoAugment (Weng, 2019), which use reinforcement learning or density matching to find optimal augmentation policies (Terrell & Scott, 1992). Additionally, gradient-based methods such as DeepAutoAugment (Zheng et al., 2022) automate policy selection without prior knowledge. These methods directly learn the augmentation policy without prior knowledge or manual selection of default transformations such as Marrie et al. Marrie et al. (2023). However, the computational complexity of these methods limits their feasibility for image classification problems with limited computational resources and time constraints. Population-based augmentation (PBA) is another promising technique that enables the simultaneous training and evaluation of multiple augmentation policies, facilitating efficient policy discovery (Ho et al., 2019). Notably, PBA has demonstrated effectiveness in discovering diverse and high-performing augmentation policies while imposing minimal computational overhead. In our study, we specifically opted for PBA due to its superior efficiency and effectiveness and further explored its potential for augmentation policy selection on a class-specific basis. It is important to note that while these methods search for policies suitable for the entire dataset, the class-dependent nature of augmentation policies has received limited attention in current research. Although the generation of class-dependent data has been studied in the context of GANs (Mirza & Osindero, 2014), to our knowledge, only a few works have explored class-dependent data augmentation (Hauberg et al., 2016; Rommel et al., 2021).

While data augmentation is commonly employed to enhance model performance, different classes within a dataset may exhibit varying sensitivities to specific transformations (Balestriero et al., 2022). This discrepancy in sensitivity becomes especially pronounced in scenarios where certain classes are subjected to per-class favoritism, leading
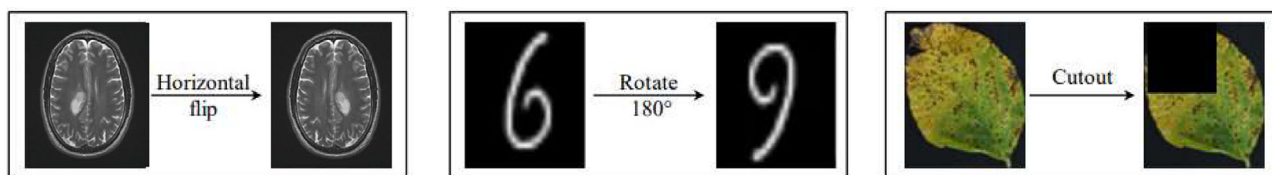
FIGURE 1 Class-specific effects of augmentations: *"horizontal flip"* distorts a brain cell image, *"vertical flip"* transforms a "6" into a "9" in MNIST, and *"cutout"* masks disease in a soybean leaf. These instances reveal that tailored strategies are essential, as not all augmentations benefit all classes.

to biased predictions and arbitrary inaccuracies on specific classes. For example, in the context of object recognition in images, using color transformations can benefit the model's ability to recognize objects such as cars or lamps, but this same augmentation strategy may have a detrimental effect on classes that are strongly defined by their color, such as apples or oranges. Similarly, applying "vertical flip" augmentation in the MNIST dataset (Deng, 2012) alters the visual representation of classes 6 and 9, as illustrated with some other examples in Figure 1.

This observation extends to plant stress classification, where distinguishing between different stress types and healthy plants can be challenging due to subtle visual differences. Several works have aimed to enhance detection accuracy using practical data augmentation technique (Cap et al., 2020; Pawara et al., 2017; Zhu et al., 2020). For instance, in cases of potassium deficiency, early identification is crucial as leaf yellowing starts from the tip of soybean leaflets. However, using cutout (DeVries & Taylor, 2017) augmentation targeting the tip of the potassium-stressed leaf might compromise the model's ability to identify potassium deficiency early on. Additionally, datasets with confounding classes (classes that are difficult to distinguish from one another due to overlapping visual characteristics or shared features) pose an additional challenge, as data augmentation can potentially worsen performance disparities among classes. Thus, applying transformations that emphasize texture or shape features to classes that are difficult to distinguish can be beneficial. Consequently, class-specific data augmentation emerges as a potent tool for enhancing ML model performance, particularly in scenarios with challenging classes.

To address the challenges posed by class-dependent invariances and to enhance classifier performance, particularly for confounding classes, we propose a novel approach that customizes augmentation strategies to capture the unique characteristics of each class. By fine-tuning a pre-trained image classification model and optimizing augmentation policies for individual classes, our class-specific approach aims to improve mean-per-class accuracy (MPCA), particularly in the context of confounding classes in the dataset. The automated process of class-specific data augmentation, driven by an

evolutionary optimization algorithm, genetic algorithm (GA; Katoch et al., 2021), selects the most effective augmentation policies for each class.

The effectiveness of our approach is demonstrated in Figure 2, where transformed images using the most and least likely augmentations for each stress class are visualized. These results highlight the efficacy of class-specific data augmentation in improving model performance. This tailored strategy strikes a balance between efficiency and effectiveness, providing a promising solution to address limitations of conventional augmentation techniques.

Specific contributions of this paper are summarized below as follows:

- We propose an effective approach based on GA to find the best set of augmentations for each class on a target dataset.
- We demonstrate the efficacy of our approach by showing that our per-class augmentations significantly improved the accuracy of the two worst-performing classes in the target dataset, increasing from 83.01% to 88.89% and 85.71% to 94.05%. Additionally, our approach significantly increased the MPCA of the dataset from 95.09% to 97.61% compared to the accuracy of the non-augmented model.

In our implementation, a well-trained classifier is used as a baseline, and it is fine-tuned for only five epochs with various sets of augmentations whose probabilities are the population created by GA. This approach significantly reduces the computational cost of searching for optimal augmentation policies while maintaining competitive performance.

## 2 | MATERIALS AND METHODS

2.1 | Dataset

The dataset used in this study is a publicly available dataset comprising 16,573 RGB images of soybean leaflets across nine distinct classes, eight different soybean stresses and healthy soybean leaflets, covering a broad range of biotic and abiotic foliar stresses (Ghosal et al., 2018). Figure 3 demonstrates the imaging setup and the nine soybean leaf
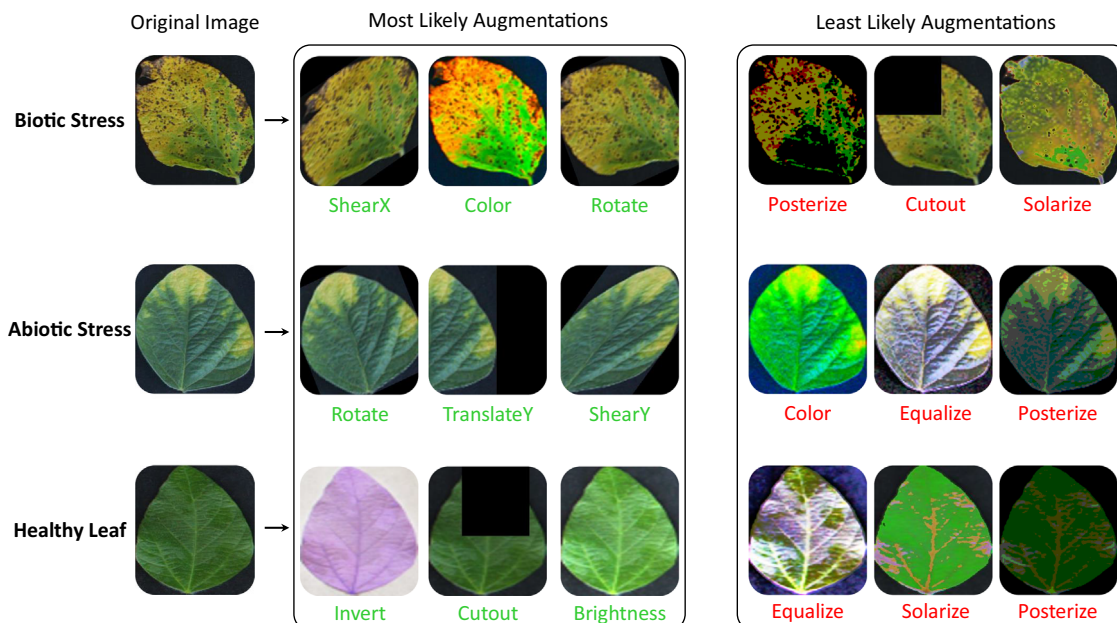
FIGURE 2 For different stress classes in the soybean stress (biotic and abiotic) dataset, we present an image from each category (left) and thin automating plant stress classification processes e corresponding image transformed using the three most likely augmentations (middle) and the three least likely augmentations (right) for that stress class, as determined by our class-specific automated data augmentation method.
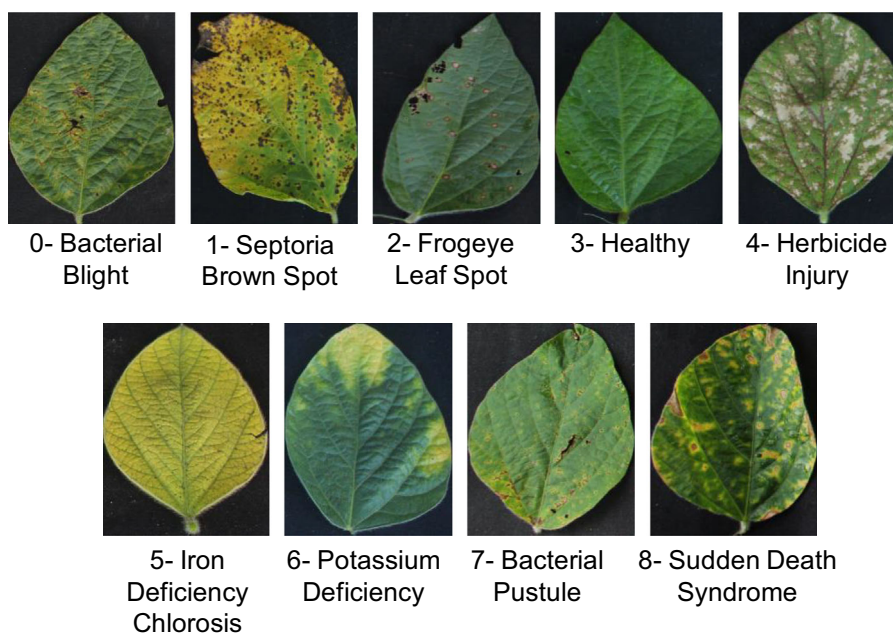

FIGURE 3 Image examples of the nine classes (healthy leaflet and eight different soybean stresses) in the dataset.

stress classes included in the analysis. The training, validation, and test datasets were composed of 13,420 (80%), 1,491 (9%), and 1,662 (11%), respectively (Table S1). More information on the dataset is available in Ghosal et al. (2018).

2.2 | Baseline model

To establish a fair and comprehensive baseline for soybean stress classification, we followed the methodology outlined in Ghosal et al. (2018), utilizing the same dataset
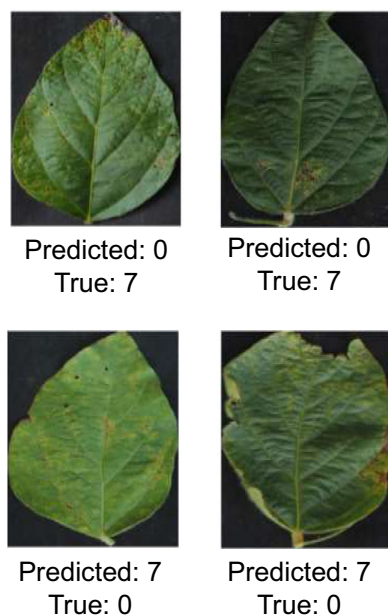
FIGURE 4 Examples of wrongly classified images by baseline model.

and model. However, to explore potential improvements, we experimented with different DL architectures to enhance the baseline accuracy. Our findings, detailed in Table S2, revealed that ResNet50 (He et al., 2016) achieved the highest accuracy (95.09%) surpassing the previously reported 94.13%, prompting its selection for further evaluation. The baseline model was trained for 350 epochs on the dataset without any data augmentations to ensure unbiased performance (Balestriero et al., 2022). During training, we utilized the categorical cross-entropy loss function, Adam optimizer with a momentum of 0.9, weight decay of 0.0001, and a batch size of 256.

Despite the promising results obtained with the baseline model, a consistent observation akin to the study's findings emerged, wherein the model encountered difficulties in accurately classifying the challenging categories of bacterial blight and bacterial pustule with per-class accuracies of 83.01% and 85.71%, respectively. Discriminating between these two stresses is challenging even for expert plant pathologists due to confounding symptoms (Hartman et al., 2015).

A few examples of misclassified images by the baseline model are provided in Figure 4, and from the figure, it is evident that these two stresses are hard to classify even for human experts. These findings highlight the need for further refinement and optimization, as an ideal classifier should excel in accurately predicting all classes, including those that pose significant challenges. To address this, our primary focus was to enhance the accuracy of the worst-performing classes, particularly targeting bacterial blight and bacterial pustule. The proposed GA-optimized automated DA algo-rithm is evaluated using our enhanced baseline model as a foundation.

## 2.3 | Genetic algorithm for optimizing data augmentations

We utilized GA, a search algorithm inspired by natural selection and genetic inheritance, to drive the evolutionary process in our study (Katoch et al., 2021). It is a method used to find the best solution to an optimization problem by exploring a population of potential solutions. Each individual in the population represents a potential solution to the problem. Through successive generations, GA iteratively explores and evolves the population, aiming to converge toward the optimal or near-optimal solution. The effectiveness of GA in achieving this goal relies on the incorporation of elitism. Elitism ensures that the best individuals from the current generation are preserved and directly transferred to the next generation without alteration. This strategy helps maintain diversity within the population while safeguarding promising solutions from premature elimination due to the randomness of genetic operations such as mutation and crossover.

In the context of our soybean leaf stress dataset, we employ GA to optimize the probability of each augmentation for the nine classes. Our data augmentation search space is composed of the standard pool of 15 transformations; ShearX/Y, Translate X/Y, Rotate, AutoContrast, Invert, Equalize, Solarize, Posterize, Contrast, Color, Brightness, Sharpness, and Cutout. These augmentations, closely aligning with those utilized in AutoAugment (Kingma & Ba, 2014), have emerged as popular choices for exploring optimal data augmentation policies in image classification tasks.

The search space for this optimization problem consists of all possible combinations of augmentation for each of the nine classes. To streamline our optimization process, we consider probabilities ranging from 0 to 1 with a step size of 0.1 for applying each augmentation to the respective class. By defining the augmentation magnitude as the mean of the possible values, we maintain consistency in the augmentation's influence. Our primary objective is to determine the most effective combination of augmentation probabilities for each class that maximizes the MPCA of our target dataset.

Our GA operators include:

- **Initialization**: Create an initial population set of probabilities ranging from 0 to 1 for each augmentation strategy.
- **Evaluation**: Assess the fitness of each augmentation strategy by evaluating its MPCA on the test dataset.
- **Selection**: Choose augmentation strategies with higher accuracy as parents for the next generation, using fitness proportionate selection or other selection strategies.
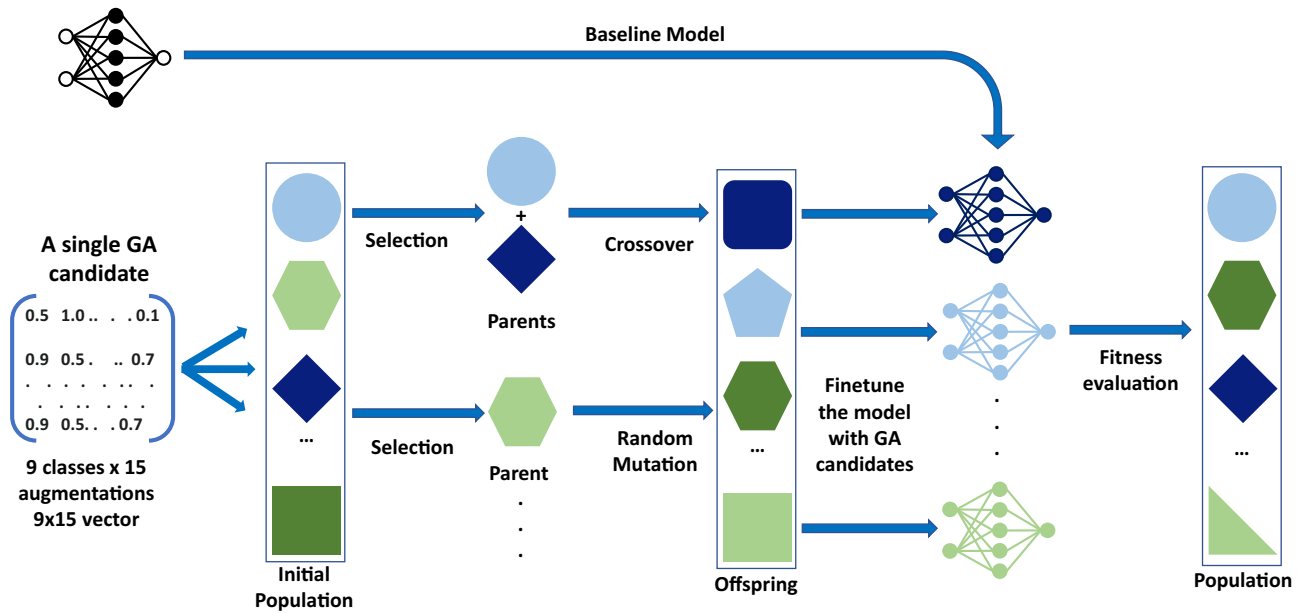
FIGURE 5   Illustration of a single generation in the genetic algorithm (GA) framework. The baseline classifier is fine-tuned with each candidate from the GA population, which represents the probabilities of augmentations for each class. These selected candidates undergo mutation and crossover operations, generating the next generation of augmentation probabilities for improved performance.

- **Crossover**: Combine probabilities of two augmentation sets to create offspring individuals with a mix of their characteristics.
- **Mutation**: Introduce random changes or modifications to the probabilities of augmentations to maintain diversity and explore new regions of the search space.

Formally, let $\mathbf{p} = (p_{ij})$ be a 9×15 matrix, where p_{ij} represents the probability of applying the j-th augmentation technique to samples from the i-th class during training. The optimization problem can be defined as follows:

$$\text{Maximize:}\quad \text{MPCA}$$

$$\text{Subject to:}$$

$$\text{Constraint:}\quad 0 \leq p_{ij} \leq 1, \quad \forall i, j$$

The objective is to maximize the MPCA and the constraints ensure that the augmentation probabilities remain within the feasible range for each decision variable. The illustration of our GA framework for a single generation is shown in Figure 5. By employing GA, we aim to effectively explore and navigate this search space, searching for the set of augmentation probabilities that leads to the highest classification accuracy on our dataset. To evaluate the performance of the classifiers, we employ commonly used evaluation met-rics, including overall accuracy, MPCA, and confusion matrix analysis:

$$\text{Mean-per-class accuracy (MPCA)} = \frac{1}{N} \sum_{i=1}^{N} \text{Accuracy}_i$$

2.4 | Fine-tuning baseline model with augmentation probabilities

Figure 6 illustrates the flowchart of the overall workflow for optimizing augmentation probabilities using GA. After generating a population of augmentation probabilities, the baseline model is fine-tuned for each augmentation probability in the population. The fine-tuning process involves applying the augmentation probabilities to the training data, evaluating the resulting classifier on the test set, and using the MPCA as the fitness score for each chromosome. Based on these fitness scores, GA performs selection, crossover, and mutation operations to generate a new population of chromosomes. This process continues iteratively until a termination criterion is met or the best solution is obtained. It is worth noting that the child networks in this study undergo a concise fine-tuning process of only five epochs, which is significantly shorter compared to other automated data augmentation strategies. We selected five epochs based on the observation of limited performance improvement beyond this point.
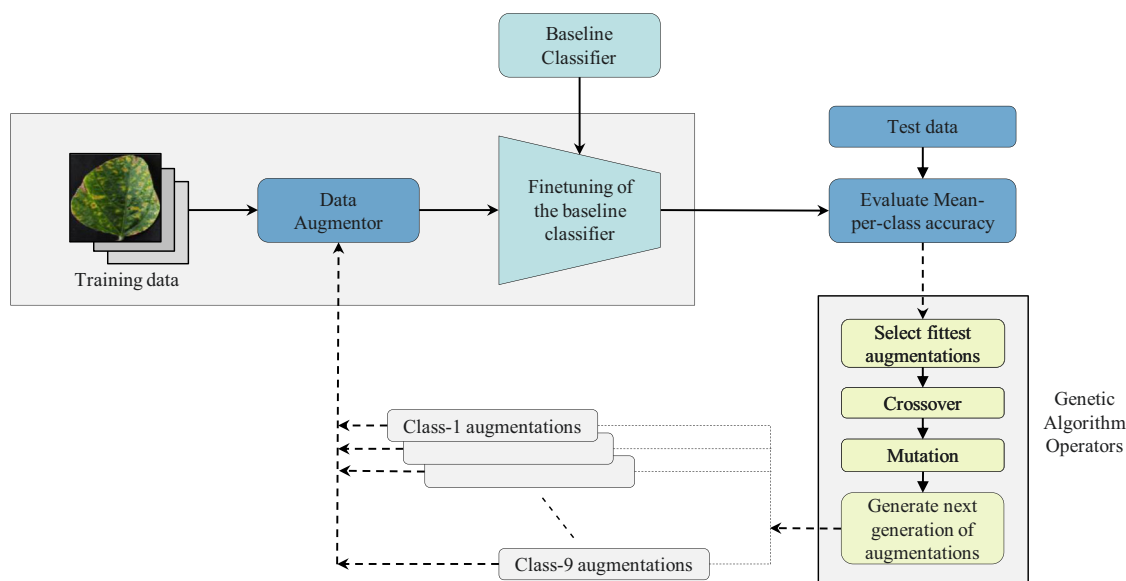
FIGURE 6 Flowchart depicting the overall workflow for optimizing augmentation probabilities using genetic algorithm.

2.5 | Implementation details

The experiments were conducted on a GPU cluster at Iowa State University, featuring four A100 NVIDIA GPUs, each equipped with 80 GB of memory. This configuration allowed us to concurrently fine-tune eight models by utilizing two models on each GPU, significantly reducing the overall processing time by running GA in parallel across 8 GPUs. On average, one generation took approximately 4.5 h to complete. As a future work, further optimization can be achieved by distributing the workload across multiple nodes, which would result in even faster processing times. Our proposed method requires less computation than traditional automated DA methods since we only fine-tune the base model using a set of augmentation probabilities for five epochs.

To implement the GA, we employed PyGAD (Gad, 2021) and configured with a maximum of 100 generations. Termination criteria were defined as either completing 100 generations or observing no improvement in fitness scores for 10 consecutive generations. Hyperparameters were optimized using the Rastrigin function, known for its challenging landscape characterized by multimodality and high oscillation (Pohlheim, 2007). A population size of 100 individuals was chosen for the GA, employing steady-state selection, random mutation, and single-point crossover to maintain diversity and explore the search space effectively.

3 | RESULTS

The primary goal of our experiments is to assess the effectiveness of automated class-specific data augmentation using a GA-based approach in improving MPCA and the accuracy of worst-performing classes. We demonstrate this by evaluating the performance of the models across each class, examining the corresponding confusion matrices, analyzing augmentations selected by GA, and the impact of the order of augmentations in classification accuracy. To ensure the robustness and generalization of our model, we conducted fivefold cross-validation on our dataset. The outcomes of this cross-validation, presented in Table S3, guided our selection of the most effective model for further investigation.

3.1 | Impact of class-specific augmentations on classification accuracy

The bar chart in Figure 7 provides a clear comparison between the baseline model and the optimized model after applying GA-based automated data augmentation. It demonstrates a substantial improvement in the MPCA, from 95.09% with the baseline model to an impressive 97.61% with the optimized model. This enhancement across all classes indicates the efficacy of employing tailored class-specific augmentations, enabling the model to better recognize and differentiate between different class characteristics, ultimately leading to more accurate classification.

Moreover, the iterative nature of the GA in selecting the most effective augmentations has significantly contributed to this improvement. Notably, the challenging classes of bacterial blight and bacterial pustule have shown substantial accuracy enhancements, with bacterial blight improving from 83.01% to 88.89%, and bacterial pustule from 85.71% to 94.05%. This underscores the importance of the GA's role in

| **The Plant Phenome Journal** OPEN ACCESS 🔓
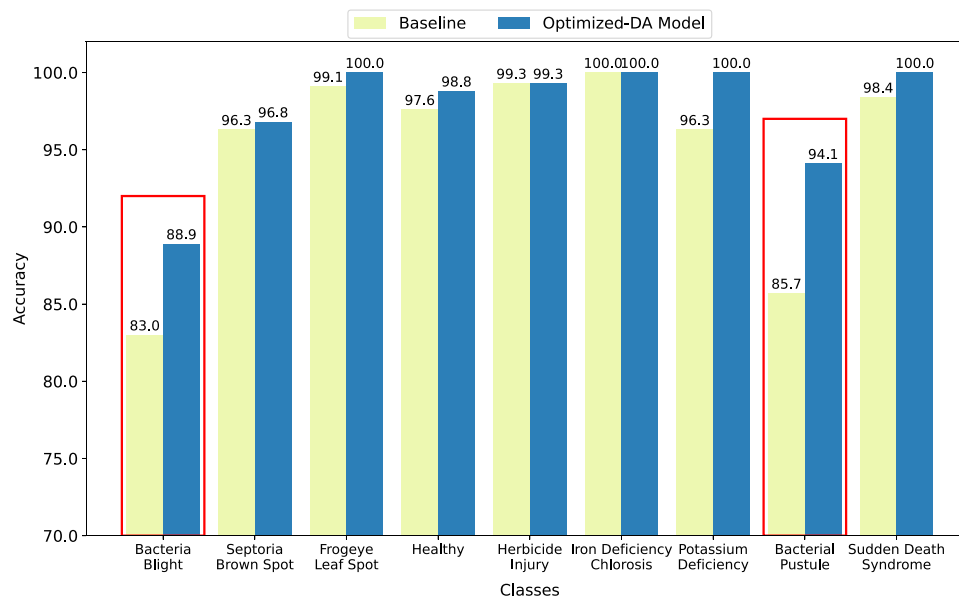
FIGURE 7 Comparison of class-wise accuracies: Bar chart comparing the accuracies of the baseline model and the optimized data augmentation (optimized-DA) model achieved through a GA-based evolutionary process. The optimized-DA model showcases remarkable improvements in accuracies for all classes, with particularly notable enhancements observed for the confounding classes—bacterial blight and bacterial pustule.

TABLE 1 Comparison with other automated augmentation methods.

Augmentation technique	Mean-per-class accuracy (%)	Sensitivity (%)	Specificity (%)
AutoAugment (ImageNet)	95.8	100	91.9
AutoAugment (CIFAR-10)	95.5	97.7	98.5
AutoAugment (SVHN)	95.6	100	93.9
RandAugment	96.2	97.7	96.3
Trivial Augment	95.9	99.0	96.9
AugMix	95.7	98.4	97.7
GA-based optimized DA (proposed method)	97.6	99.2	97.0

Abbreviations: GA, genetic algorithm; DA, data augmentation.

identifying and implementing augmentations specifically tailored to address the unique challenges posed by these classes. Overall, these results demonstrate the effectiveness of our class-specific DA approach in overcoming class-specific challenges and significantly improving classification accuracy.

In a comprehensive comparison with other automated augmentation methods on the soybean disease dataset, as detailed in Table 1, our method notably surpasses all others in terms of accuracy. This highlights the effectiveness of our proposed approach. Importantly, our method achieves superior accuracy while significantly reducing computation requirements by only fine-tuning the baseline model for five epochs, without training any augmentation policy from scratch. This streamlined approach not only enhances accuracy but also optimizes computational resources, making it a practical solution for real-world applications.

3.2 | Impact of class-specific augmentations on misclassifications

To assess the performance of the models on misclassifications, we analyzed the confusion matrices of the baseline model and the augmented model (Figure 8). As mentioned earlier, the baseline model struggled particularly with predicting bacterial blight (class 0) and bacterial pustule (class 7), frequently misclassifying them interchangeably (Hartman et al., 2015). However, the optimized model exhibited a noticeable reduction in misclassifications for these challenging classes. By tailoring augmentations to each class, the GA automatically selects augmentations that help distinguish these classes from each other. Consequently, the optimized model showed improved per-class accuracies, suggesting that our class-specific DA techniques effectively addressed the
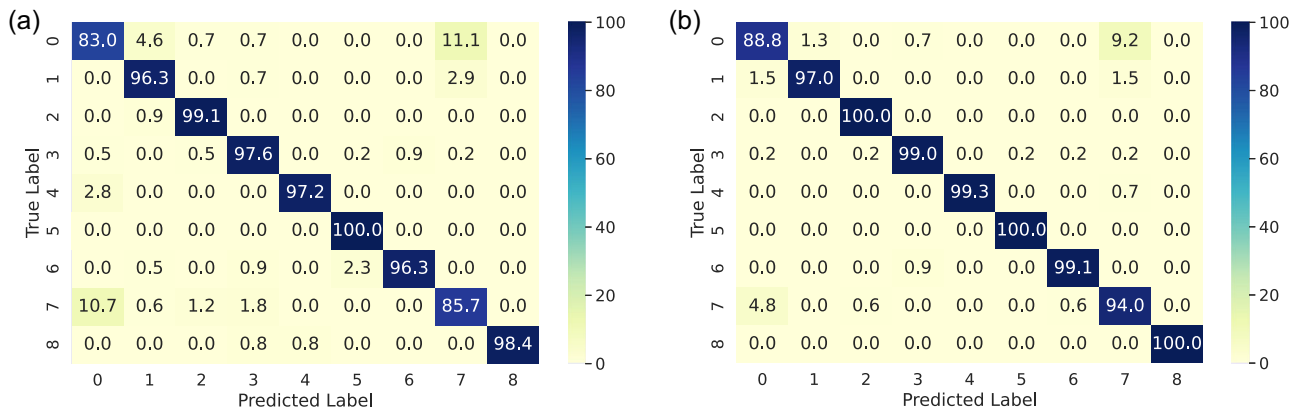
FIGURE 8 Comparison of classification accuracy confusion matrices: (a) Baseline model; (b) optimized data augmentation (optimized-DA) model. The augmented model demonstrates improved per-class accuracies, as evident from the reduction in misclassifications illustrated in the confusion matrices. Particularly, in the case of class 0 (bacterial blight) and class 7 (bacterial pustule), the misclassifications have significantly reduced, highlighting the effectiveness of our approach in addressing the challenges associated with these classes.
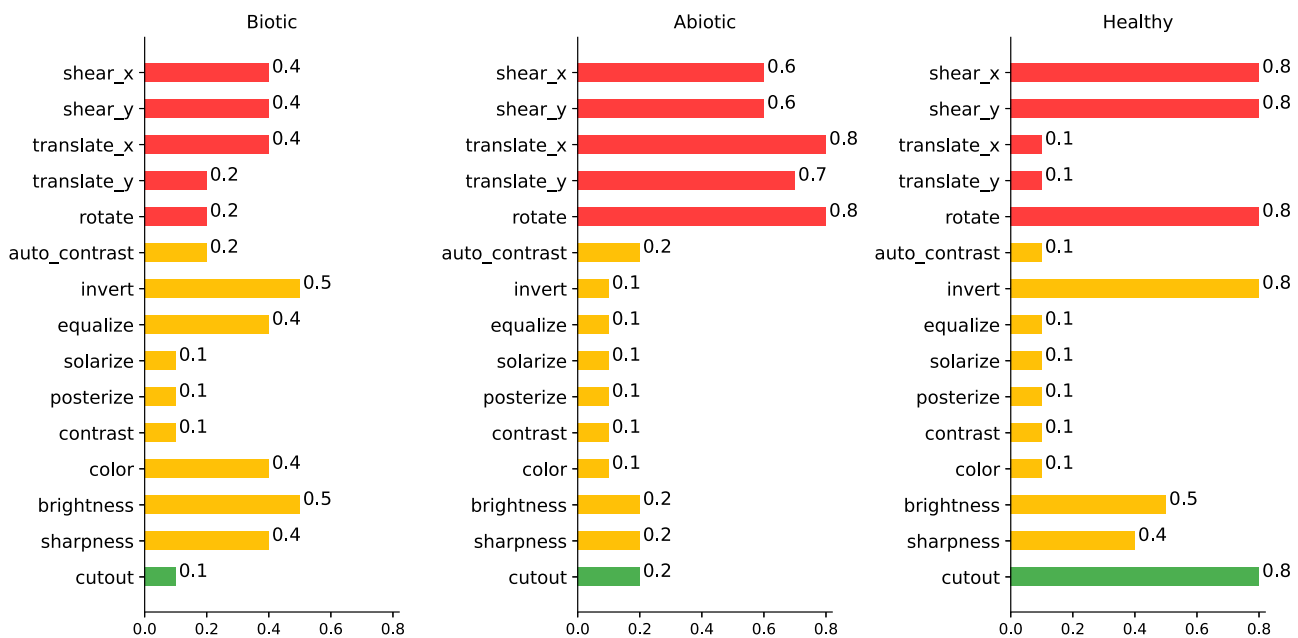
FIGURE 9 Optimized augmentation policies for different stress conditions (biotic and abiotic) and healthy leaves. Augmentation techniques are categorized into three groups: Geometry-based augmentations (red), color-based augmentations (yellow), and the cutout (green).

baseline model's limitations. These enhancements validate the effectiveness of our approach in improving classification performance, especially for the most challenging classes.

3.3 | Comparison of optimized augmentations on different stresses

In our analysis of the optimized augmentations, we aimed to understand their impact on different stress conditions, including biotic and abiotic stresses, as well as healthy leaves.

The optimized augmentation policies, depicted in Figure 9, shows the preferences for specific augmentation types across these classes (Gull et al., 2019).

For biotic stress classes, we observed consideration for both color-based and geometry-based augmentations. However, specific color augmentations such as solarize, posterize, and invert were not favored due to their limited relevance to disease-related visual cues in this context. Conversely, in the case of abiotic stress classes, geometry-based augmentations were preferred, with shear, translation, and rotation being prominent choices. Additionally, augmentations like

TABLE 2 Accuracy comparison of different order of augmentations.

Order of augmentations	Accuracy (%)
Geometry → Color → Cutout	97.6
Geometry → Cutout → Color	96.7
Color → Geometry → Cutout	96.5
Cutout → Geometry → Color	96.4
Color → Cutout → Geometry	96.4
Cutout → Color → Geometry	95.9

sharpness and autocontrast were selected for their effectiveness in capturing the structural changes associated with abiotic stressors.

In contrast, for the healthy class, geometry-based augmentations were predominantly chosen, with brightness, sharpness, and autocontrast selected to enhance the natural appearance of healthy leaves. Interestingly, the cutout augmentation was exclusively chosen by the healthy class, while being avoided by other stress classes. This suggests that cutout augmentation, which masks specific regions in the images, may inadvertently remove relevant disease-related information for other classes. Overall, the analysis highlights the importance of selecting appropriate augmentations tailored to each stress class to improve the accuracy of soybean stress classification.

3.4 | Does the order of augmentations matter?

To explore the impact of augmentation order on model performance, separate GA runs were conducted for each proposed augmentation sequence, categorized into three: (i) Geometry—includes augmentations that modify the geometric properties of the images, such as shearing and rotation; (ii) Color—comprises augmentations that manipulate the color and contrast characteristics of the images; and (iii) Cutout. Table 2 provides a summary of the results obtained from these runs.

We observed that the augmentation order has a slight influence on the model's performance. The highest accuracy of 97.6% was achieved when the augmentations were applied in the order of Geometry, followed by Color, and then Cutout (Perez & Wang, 2017). This implies that initiating the augmentation process with geometric transformations, followed by color manipulations, and concluding with cutout techniques can lead to superior accuracy in the context of our dataset. It is important to note that these findings are specific to our dataset, and results may vary for different datasets. Despite a small relative change in accuracies, the overall performance remained consistently high across all orders,

indicating that the choice of augmentation order may not be critical in achieving strong results, and further research is needed in other datasets. These findings suggest that while the order of augmentations may have a marginal impact on the model's performance, the selection and combination of augmentation techniques play a more significant role in improving accuracy.

4 | DISCUSSION

DL models often struggle to achieve consistent high performance across all classes within a dataset, despite achieving high overall accuracy. Data imbalances and lack of diversity in the training data are among the key reasons for this phenomenon. Data augmentation, which aims to enhance model performance and mitigate the challenges imposed by data imbalances and diversity, has emerged as an effective approach to address these issues. In this study, we demonstrate that by tailoring augmentations specific to each class in a dataset, these limitations can be effectively mitigated.

To achieve this, we deployed tailored augmentations for each class in our soybean disease dataset using GA-based optimization. We fine-tuned a well-trained baseline model for each data augmentation policy generated by the GA. Through comprehensive evaluation of MPCA and confusion matrices, we observed significant improvements in the accuracy of each class in the dataset. Notably, the accuracy of confounding classes, such as bacterial blight and bacterial pustule, has also been substantially improved.

The key mechanism behind our approach lies in the utilization of GA to automatically select augmentations tailored to each class. The GA iteratively explores the augmentation space and identifies the most effective augmentations that maximize the MPCA. By fine-tuning the baseline model with these augmentations for a limited number of epochs, we efficiently enhance the model's ability to distinguish between different classes and improve overall classification performance. This adaptive and iterative approach bears resemblance to boosting techniques in ML (Tanha et al., 2020). Just as boosting algorithms iteratively train weak learners to create a strong ensemble model that excels in classifying difficult instances, our method iteratively refines the baseline model by selecting augmentations tailored to address the challenges posed by specific classes. The GA's exploration of the augmentation space parallels the boosting process of focusing on misclassified instances in successive iterations, ultimately leading to improved classification performance.

Moreover, our approach offers two distinct advantages over existing techniques (Cubuk et al., 2018; Ho et al., 2019; Lim et al., 2019; Marrie et al., 2023). First, unlike previous methods that optimize augmentations for the entire dataset, we tailor augmentations specific to each class in

the dataset. This class-specific nature allows our approach to address the unique characteristics and challenges associated with individual classes, resulting in improved performance across all classes. Second, our method significantly reduces computation requirements by only fine-tuning the last layer of the model for a limited number of epochs for each augmentation policy generated by the GA. This streamlined approach not only enhances accuracy but also optimizes computational resources, making it a practical solution for real-world applications

Furthermore, our analysis of the augmentations selected by the optimized model reveals interesting insights into the preferences of specific stresses (biotic, abiotic, and healthy) for particular augmentation types. This verifies that classes within a dataset can indeed have different preferences for augmentations, highlighting the importance of class-specific augmentation strategies.

Additionally, we investigated the effect of the order in which augmentations are applied on model performance. Our results indicate that while the augmentation order may have a slight influence on performance, the selection and combination of augmentation techniques play a more significant role in improving accuracy. Initiating the augmentation process with geometric transformations, followed by color manipulations, yielded superior accuracy in our dataset.

Overall, our study underscores the effectiveness of tailored class-specific data augmentations in enhancing DL model performance for soybean stress classification. By addressing class-specific challenges and optimizing the augmentation process, our approach offers a promising solution for accurate disease diagnosis and management in agricultural applications. Future research directions may involve exploring the application of our method to different crops and stress conditions, as well as investigating the integration of advanced ML techniques for further performance enhancement.

5 | CONCLUSION

This study demonstrates the efficacy of a GA-based approach in identifying class-specific augmentations to improve plant stress classification accuracy. By fine-tuning a baseline model with tailored augmentations, we achieved a notable increase in MPCA, with the optimized model achieving an impressive average per-class accuracy of 97.61%, surpassing the performance of existing automated augmentation methods. Particularly, previously challenging classes such as bacterial blight and bacterial pustule showed significant accuracy enhancements, with bacterial blight accuracy increasing from 83.01% to 88.89% and bacterial pustule accuracy jumping from 85.71% to 94.05%. These improvements highlight the effectiveness of our approach in addressing class-specific challenges.

The findings of this study underscore the importance of tailored augmentation strategies for individual classes in plant stress classification tasks. Leveraging GA optimization, we showcased significant improvements in accuracy, providing valuable insights for the development of class-specific augmentation techniques. These results have implications beyond soybean disease classification, offering guidance for similar classification tasks in agriculture and other domains.

AUTHOR CONTRIBUTIONS
Nasla Saleem: Data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft. **Aditya Balu**: Project administration; resources; supervision; writing—review and editing. **Talukder Jubery**: Supervision; validation; writing—review and editing. **Arti Singh**: Supervision; writing—review and editing. **Asheesh Singh**: Supervision; writing—review and editing. **Soumik Sarkar**: Supervision; writing—review and editing. **Baskar Ganapathysubramanian**: Conceptualization; funding acquisition; methodology; project administration; resources; supervision; validation; writing—review and editing.

CONFLICT OF INTEREST STATEMENT
The authors declare that there is no conflict of interest regarding the publication of this article.

DATA AVAILABILITY STATEMENT
The data used for the creation of this manuscript are available at https://github.com/nasla-96/Class-specific-Data-Augmentation-for-Plant-Stress-Classification/tree/master

ORCID
Arti Singh https://orcid.org/0000-0001-6191-9238
Asheesh K. Singh https://orcid.org/0000-0002-7522-037X
Baskar Ganapathysubramanian https://orcid.org/0000-0002-8931-4852

REFERENCES
Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., & Alrahamneh, Z. (2011). Fast and accurate detection and classification of plant

diseases. *International Journal of Computer Applications*, *17*(1), 31–38.

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, *19*(1), 52–61.

Atefi, A., Ge, Y., Pitla, S., & Schnable, J. (2021). Robotic technologies for high-throughput plant phenotyping: Contemporary reviews and future perspectives. *Frontiers in Plant Science*, *12*, 611940.

Balestriero, R., Bottou, L., & LeCun, Y. (2022). The effects of regularization and data augmentation are class dependent. *Advances in Neural Information Processing Systems*, *35*, 37878–37891.

Cap, Q. H., Uga, H., Kagiwada, S., & Iyatomi, H. (2020). Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering*, *19*(2), 1258–1267.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501* https://doi.org/10.48550/arXiv.1805.09501

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, *29*(6), 141–142.

DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*.

Feng, L., Chen, S., Zhang, C., Zhang, Y., & He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, *182*, 106033.

Gad, A. F. (2021). Pygad: An intuitive genetic algorithm python library. arXiv: 2106.06158 [cs.NE].

Gao, T., Emadi, H., Saha, H., Zhang, J., Lofquist, A., Singh, A., Ganapathysubramanian, B., Sarkar, S., Singh, A. K., & Bhattacharya, S. (2018). A novel multirobot system for plant phenotyping. *Robotics*, *7*(4), 61.

Gao, Z., Luo, Z., Zhang, W., Lv, Z., & Xu, Y. (2020). Deep learning application in plant stress imaging: A review. *AgriEngineering*, *2*(3), 29.

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, *115*(18), 4613–4618.

Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., & Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics*, *2*(3), 156–183.

Gonzalez Guzman, M., Cellini, F., Fotopoulos, V., Balestrini, R., & Arbona, V. (2022). New approaches to improve crop tolerance to biotic and abiotic stresses. *Physiologia Plantarum*, *174*(1), e13547.

Gull, A., Lone, A. A., & Wani, N. U. I. (2019). Biotic and abiotic stresses in plants. In *Abiotic and biotic stress in plants* (pp. 1–19). IntechOpen.

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., Singh, A. K., & Ganapathysubramanian, B. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics*, *2021*, 9840192.

Hartman, G. L., Rupe, J. C., Sikora, E. J., Domier, L. L., Davis, J. A., & Steffey, K. L. (2015). *Compendium of soybean diseases and pests*. American Phytopathological Society.

Hauberg, S., Freifeld, O., Larsen, A. B. L., Fisher, J., & Hansen, L. (2016). Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In *Proceed-ings of the Artificial Intelligence and Statistics* (pp. 342–350). PMLR.

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* (pp. 770–778). IEEE, Piscataway, NJ.

He, X., Zhao, K., & Chu, X. (2021). Automl: A survey of the state-of-the-art. *Knowledge-Based Systems*, *212*, 106622.

Herr, A. W., Adak, A., Carroll, M. E., Elango, D., Kar, S., Li, C., Jones, S. E., Carter, A. H., Murray, S. C., Paterson, A., & Sankaran, S. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, *63*(4), 1722–1749.

Ho, D., Liang, E., Chen, X., Stoica, I., & Abbeel, P. (2019). Population based augmentation: Efficient learning of augmentation policy schedules. In *International Conference on Machine Learning* (pp. 2731–2741). PMLR.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90.

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications*, *80*, 8091–8126.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations. arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (Vol. 25). Curran Associates.

Lim, S., Kim, I., Kim, T., Kim, C., & Kim, S. (2019). Fast autoaugment. In *Advances in neural information processing systems* (Vol. 32). Curran Associates.

Marrie, J., Arbel, M., Larlus, D., & Mairal, J. (2023). SLACK: Stable learning of augmentations with cold-start and KL regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 24306–24314). IEEE.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Mosa, K. A., Ismail, A., & Helmy, M. (2017). Introduction to plant stresses. In *Plant stress tolerance: An integrated omics approach* (pp. 1–19). Springer.

Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., & Ganapathysubramanian, B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, *14*, 1–13.

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics*, *2019*, 5809404.

Pawara, P., Okafor, E., Schomaker, L., & Wiering, M. (2017). Data augmentation for plant classification. In *Advanced Concepts for Intelligent Vision Systems* (pp. 615–626). Springer.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Pieruschka, R., & Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenomics*, *2019*, 7507131. https://spj.science.org/doi/abs/10.34133/2019/7507131

Pohlheim, H. (2007). Examples of objective functions. *Retrieved*, 4(10), 2012.

Rebuffi, S.-A., Gowal, S., Calian, D. A., Stimberg, F., Wiles, O., & Mann, T. A. (2021). Data augmentation can improve robustness. *Advances in Neural Information Processing Systems*, 34, 29935–29948.

Rommel, C., Moreau, T., Paillard, J., & Gramfort, A. (2021). CADDA: Class-wise automatic differentiable data augmentation for EEG signals. *arXiv preprint arXiv:2106.13695*.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, 72(1), 1–13.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, 21(2), 110–124.

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends in Plant Science*, 23(10), 883–898.

Singh, D. P., Singh, A. K., & Singh, A. (2021). *Plant breeding and cultivar development*. Academic Press.

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: An experimental review. *Journal of Big Data*, 7, 1–47.

Taylor, L., & Nitschke, G. (2018a). Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1542–1547). IEEE. doi: https://doi.org/10.1109/SSCI.2018.8628742

Terrell, G. R., & Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, 20, 1236–1265.

Van Dyk, D. A., & Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1), 1–50.

Weng, L. (2019). From gan to wgan. *arXiv preprint arXiv:1904.08994*.

Xu, B., Fan, J., Chao, J., Arsenijevic, N., Werle, R., & Zhang, Z. (2023). Instance segmentation method for weed detection using UAV imagery in soybean fields. *Computers and Electronics in Agriculture*, 211, 107994.

Zhang, J., Naik, H. S., Assefa, T., Sarkar, S., Reddy, R. C., Singh, A., Ganapathysubramanian, B., & Singh, A. K. (2017). Computer vision and machine learning for robust phenotyping in genome-wide studies. *Scientific Reports*, 7(1), 44048.

Zheng, Y., Zhang, Z., Yan, S., & Zhang, M. (2022). Deep autoaugmentation. *Proc. ICLR*, 1(3), 6.

Zhu, F., He, M., & Zheng, Z. (2020). Data augmentation using improved cdcgan for plant vigor rating. *Computers and Electronics in Agriculture*, 175, 105603.

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., & Le, Q. V. (2020). Learning data augmentation strategies for object detection. In *European Conference on Computer Vision* (pp. 566–583). Springer.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.