Received: 30 November 2023 | Revised: 2 February 2024 | Accepted: 18 February 2024

IET Image Processing

DOI: 10.1049/ipr2.13069

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Enhanced classification performance through GauGAN-based data augmentation for tomato leaf images

Seung-Beom Cho[1] | **Yu Cheng**[2] | **Sanghun Sul**[2]

[1]School of Mechanical Engineering, Sungkyunkwan University, Suwon, Republic of Korea

[2]School of Service Design Convergence, Sungkyunkwan University, Suwon, Republic of Korea

Correspondence
Sanghun Sul, School of Service Design Convergence, Sungkyunkwan University, Suwon, Republic of Korea.
Email: sanghunsul@skku.edu

Abstract

This study investigated a data augmentation method for plant disease classification and early diagnosis based on a generative adversarial neural network (GAN). In the development of classification models using deep learning, data imbalance is a primary factor that reduces classification performance. To address this issue, tomato disease images from the public dataset PlantVillage were used to evaluate the performance of the GauGAN algorithm. The images generated by the proposed GauGAN model were used to train a MobileNet-based classification model and compared with methods trained with conventional data augmentation techniques and cut-mix and mix-up algorithms. The experimental results demonstrate that based on F1-scores, GauGAN-based data augmentation outperformed conventional methods by more than 10%. In addition, after the model was retrained on data collected in the field, it efficiently generated various disease images. The evaluation results from those images also revealed a data augmentation effect of about 10% compared with traditional augmentation techniques.

1 | INTRODUCTION

Tomatoes are one of the world's most produced and consumed vegetables, and their production is increasing year after year. Crop diseases have a significant impact on tomato yield and quality, and early detection and diagnosis of these diseases is critical. PlantVillage is a popular dataset for research in this area, and there is also active research on field data with various characteristics. The PlantVillage dataset primarily consists of lab-scale images with a consistent background distribution. While this dataset is balanced across ten disease categories, collecting large-scale disease images in the field is time-consuming and labor-intensive, potentially leading to data imbalances. The resulting data imbalance can have a significant impact on the ability of a model to generalize [1]. As a result, robust data augmentation techniques are desperately needed to address these imbalances and improve the model's performance.

Traditional data augmentation techniques are based on image transformations such as rotation, flip, translation, and color manipulation, which have the benefit of being quick and simple to implement. However, those techniques offer limited improvements to the generalization performance of a model due to the loss of essential image information and their inability to accurately reflect an image's visual characteristics [2, 3].

Progress in the field of Generative Adversarial Networks (GANs) provides an opportunity for addressing this challenge. Since their introduction in 2014 by Ian Goodfellow and colleagues, GANs have made significant advances in a variety of fields, including image analysis [4]. For example, they can create realistic visuals, transform the visual quality of images, and improve image resolution [5, 6]. Recently, the application of a GAN algorithm has been proposed as an extension of the public PlantVillage dataset to improve plant disease classification and early diagnosis [7–9].

However, not all GAN models are appropriate for all tasks. For example, CycleGAN is a representative GAN algorithm for data augmentation that performs well for style transfer. Through its style transformation, CycleGAN aims to make the data distribution of the target image similar to the data distribution of the style image [10, 11]. However, the transfer does not fully preserve the original structure, resulting in image information loss, which isn't ideal for our purposes [10].

In this context, we focus on the GauGAN (Gaussian-GAN) algorithm, developed by Park et al [12]. This algorithm functions

as a conditional generative model, adept at generating realistic images while preserving the structural integrity of the subject. For our study, this is critical in maintaining the geometric details of leaf diseases.

The main contribution of this paper is the innovative application and modification of the GauGAN algorithm for data augmentation in plant disease classification, particularly for tomato crops. This approach addresses the challenges of data imbalance in the PlantVillage dataset and overcomes the limitations of traditional augmentation methods. By implementing a standardized procedure for segmentation maps and utilizing a one-hot encoded map as input, we have tailored the GauGAN algorithm for a more sophisticated data augmentation process. Our methodology not only generates high-quality images but also preserves the essential geometrical features of leaf disease, pivotal for accurate classification.

The remaining sections of this paper were organized as follows. In Section 2, prior work was reviewed with a focus on existing techniques, GAN-based data augmentation, and laboratory versus field datasets. In Section 3, our proposed methodology was described. In Section 4, the experimental results were analyzed and compared with the results of existing methodologies. Section 5 presented conclusions and directions for future research.

2 | RELATED WORKS

In this study, we focus on improving how well an algorithm classifies leaf disease by augmenting a limited dataset with additional data. Insufficient datasets, in terms of both size and variety, pose the greatest challenge to the use of convolutional neural network (CNN)-based deep-learning algorithms for plant disease classification. However, the process of data collection can be significantly affected by external factors such as weather and seasonal variations, and annotating images is typically laborious and time-consuming. To address those data imbalances, various data augmentation techniques have been developed, such as AugMix [13], population-based augmentation [14], fast AutoAugment [15], RandAugment [16], and CutMix [17]. In this section, we summarize the research on traditional and GAN-based data augmentation techniques as they are applied in a variety of fields.

2.1 | Data augmentation for plant disease data

Mohanty et al. [18] evaluated the classical network models AlexNet and GoogLeNet using a public database of 54,306 images acquired under controlled conditions to identify 14 crop species and 26 diseases. The classification models demonstrated an accuracy of 99.35%; however, when tested on a set of images captured under conditions different from those used for training, the accuracy of the models decreased greatly. Fuentes et al. [19] presented a robust deep learning–based detector for real-time recognition of tomato diseases and pests on a variety of field-collected images. The data collected in the field included various backgrounds and lighting conditions and objects of varying sizes.

In addition to disease classification, data augmentation techniques are used in numerous other fields. Su et al. [20] proposed a novel framework for data augmentation based on a random image cropping and patching algorithm. They evaluated their method using two datasets from two farms and discussed its limitations when a large amount of training data is readily available. In terms of accuracy and average intersection over union values, their data augmentation method outperformed conventional methods.

Enkvetchakul et al. [21] proposed a system for detecting plant leaf diseases using the MobileNetV2 and NasNetMobile architectures for a deep CNN. Due to its small model size, their CNN architecture can be designed for smartphones, and they investigated a mixed training technique combining online and offline learning about two plant leaf diseases. During their experiments, the NasNetMobile architecture demonstrated the highest precision, and the combination of offline learning and data augmentation techniques modified that precision.

Lamba et al. [22] trained a ResNet-based model on approximately 2,215 rice leaf images from the PlantVillage dataset. Seven data augmentation techniques were used to increase the number of images to 12,684, and they were then used to identify various plant species and disease categories. On the test dataset, the trained model achieved 99.55% accuracy and 0.0029 validation loss, demonstrating the effectiveness of their approach.

2.2 | GAN-based data augmentation

As one GAN-based data augmentation technique to address data imbalance, style transfer–based GANs use image style transformation techniques to augment data, transferring the original image's style to another image. Cap et al. [23] proposed LeafGAN, an improved version of CycleGAN, to transform images of healthy cucumber leaves into images of disease. Using a background segmentation mask generated by a weakly supervised learning technique, LeafGAN applied the transformation only to the leaf region. As a result, it preserved the background information of the original image in the transferred images, potentially improving image quality.

GUO et al. [24] proposed a maize disease image generation algorithm using an enhanced CycleGAN with attention mechanisms and feature loss functions, demonstrating its ability to produce high-quality diseased maize leaf images and improve classification accuracy rates to over 91%. ZHANG et al. [25] introduced a novel high-quality image augmentation method using dual GANs for generating enhanced rice leaf disease images, which notably improves the recognition accuracy in deep learning models, addressing challenges in smart agriculture. Min et al. [26] proposed a data augmentation method based on an image-to-image translation model that provided supplements for insufficient images of diseased leaves. Their method for augmentation performed a translation between images of

FIGURE 1 Limitation of the Pix2pix algorithm: Low-quality images were generated due to overfitting in the test set data.

healthy and diseased leaves and used an attention mechanism to generate images with more distinct disease textures. That enhancement generated more plausible diseased leaf images than previous methods and verified that their data augmentation method could enhance the performance of classification models for early plant disease diagnosis.

Region emphasis based GANs augment data by focusing on and enhancing specific regions within an image by emphasizing specific objects or features. Sun et al. [27] focused on object-level texture style transfer in fashion images and proposed a two-stage unsupervised approach called TsrNet to reduce the time and cost of data preparation. Various texture transitions were performed using the multiscale module and gradient-based texture structure loss (GTS) were used with Mask-GAN to segment texture regions and with CycleGAN to exchange textures.

Zhang proposed dual GANs to generate high-resolution rice blast images: WGAN-GP to generate rice disease images and Optimized-Real-ESRGAN to improve the resolution of the images. Those experimental results showed that the accuracy of ResNet-18 and VGG-11 improved by 4.57% and 4.1%, respectively [28].

Outside the domain of plant disease data, Strelcenia et al. [29] conducted research on a system to identify credit card fraud transactions. They proposed K-CGAN, a variant of the conditional-GAN architecture (CGAN) that uses Kilberg divergence as a custom loss function, to augment data and improve the identification of fraudulent transactions. They compared their system with the SMOTE, B-SMOTE, and ADASYN techniques and demonstrated superior performance in terms of recall, Fl-score, accuracy, and precision.

3 | MATERIALS AND METHODS

3.1 | Challenges

To overcome the shortcomings of the DCGAN model [30, 31] in, which uses random latent vectors to generate high-quality images of shapes and colors, CycleGAN-based style transfer has gained popularity. However, the CycleGAN model is built on the pix2pix algorithm [32], which requires a significant amount of data. An insufficient amount of training data significantly lowers the image quality when predicting the validation data, as illustrated in Figure 1. Furthermore, because it is a one-to-one correspondence style-transfer model, the pix2pix algorithm has the disadvantage that input images must be paired with generated images. As a result, although there is no problem generating normal images, the number of disease images that pix2pix can generate is limited to the number of disease images in the training dataset. In other words, pix2pix-based data augmentation cannot maximize the effectiveness of already imbalanced datasets.

In this context, we aim to develop a data augmentation technique that can maximize the performance of tomato blight classification models by generating images of different shapes while maintaining the boundaries of leaf regions. This is the first attempt to investigate the suitability of the GauGAN algorithm for tomato leaf diseases. This research will contribute to the advancement of GauGAN's suitability and applicability.

Figure 2 depicts the structure of the overall framework proposed in this paper. The framework contains a step for constructing training images through image preprocessing and a step for constructing the network architecture of the training models based on the Gaussian distribution. We describe the image preprocessing steps used to create images that will be used as input data for the model in detail, along with the SPADE, generator, and discriminator network structures, which are the central components of the network architecture for the proposed model.

3.2 | Dataset

In this study, research was conducted based on two distinct datasets: the PlantVillage dataset [33] collected in a laboratory environment, and a Field dataset gathered directly from agricultural fields. The PlantVillage data were employed to train

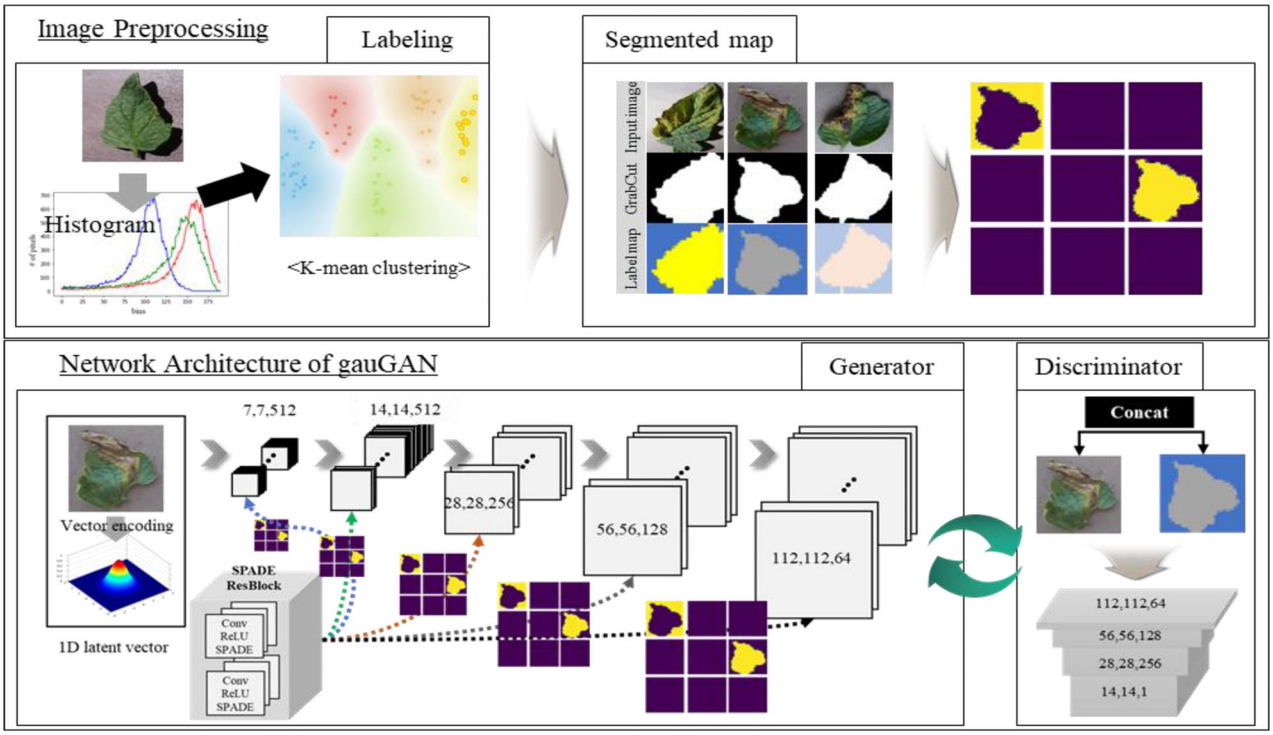

FIGURE 2 Image preprocessing and the basic structure of the GauGAN network. Image preprocessing and detailed architecture of generator-discriminator for segmentation map generation.

TABLE 1 Types and quantities of data used in experiment.

	PlantVillage					
	Original	**Tuning**	**Val_1**	**Val_2**	**Val_3**	**Field data**
Healthy	1,591	636	1,591	1,591	1,591	776
Early blight	1,000	400	200	1,000	1,000	97
Leaf mold	952	381	952	200	952	
Leaf spot	1,771	708	1,771	1,771	200	

FIGURE 3 An example of a PlantVillage dataset and field dataset.

and fine-tune the GauGAN algorithm, specifically focusing on the optimization of its loss function. This optimized GauGAN algorithm was then tested and validated using the field-collected dataset to evaluate its practical applicability and effectiveness in real-world agricultural settings. Table 1 shows the datasets and image quantities used in this study, and Figure 3 depicts the field-collected data that we gathered as photographs at a tomato farm.

For this, we conducted a parameter study to develop a GauGAN model suitable for data augmentation using only 40% of the original data. The disease states applied in this study were limited to healthy, early blight, leaf mold, and leaf spot. Subsequently, we intentionally reduced the data for distinct diseases to create imbalanced datasets Val_1, Val_2, and Val_3. Val_1 targets early blight, Val_2 targets leaf mold, and Val_3 targets leaf spot. The purpose of this evaluation was to assess the robustness against imbalanced datasets, regardless of the imbalance ratio. We evaluated the effects of data augmentation on early blight based on these datasets to determine the effectiveness of this method in addressing and counteracting data imbalance.

3.3 | Image data preprocessing for GauGAN training

This section will provide a detailed explanation of the preprocessing steps necessary for image data before training the GauGAN model. The input data for the GauGAN model must

include training images and label information, so it requires a suitable preprocessing step. The label information for each image characteristic was classified using k-means clustering [34] after the original image's histogram has been computed. The histogram is useful for identifying the overall brightness or color distribution of an image. After the images were embedded with a pre-trained CNN model, K-means clustering was applied to the input images, which allowed more accurate label data to be obtained.

That value is then assigned to a segmented binary image of the tomato, where the background is represented as 0 and the leaf area as 1. Colors are assigned based on the assigned label data, and similar or duplicate colors are not set for the background and leaf areas. Colors that were similar were found to contribute to overfitting when applied to the GauGAN model, so we used colors that were easily distinguishable. After that, these assigned color images were converted into a one-hot label map, with each pixel represented as a distinct label, a format commonly used for precise image segmentation. In this configuration, the position of a pixel that corresponded to its label was marked as 1, while other positions were marked as 0. A representation like this provides a clear view of the label associated with each individual pixel. As such, to improve the generalization ability of the Gaussian model, we preprocessed the input image to capture detailed label information on a pixel-by-pixel basis.

3.4 | Network architecture of GauGAN

3.4.1 | Utilizing variational encoder in GauGAN: Extracting high-dimensional features

The role and structure of the variational encoder in GauGAN will be analyzed in this section, explaining how it extracts and utilizes high-dimensional features of images. It emphasizes the significance of the variational encoder as a crucial element in the GauGAN architecture. Variational Autoencoders inspired GauGAN, which uses a variational formulation in which an encoder learns the mean and variance of a normal (Gaussian) distribution from cue images. The encoder is better equipped to capture the inherent randomness and variations present in real-world imagery by learning the mean and variance of a Gaussian distribution from cue images. This stochastic approach ensures that the generated images are not just deterministic outputs but can display variations similar to how nature or real-world scenarios would present them. The GauGAN can also ensure that the latent space where the image attributes are embedded is smooth and continuous by leveraging this Gaussian distribution. This continuous latent space is critical for maintaining the fidelity of the generated images, especially when the input is slightly modified. The variational encoder ensures that as users interact with GauGAN, making small modifications to the segmentation maps or input cues, the generated outputs reflect these changes in a coherent and visually consistent manner.

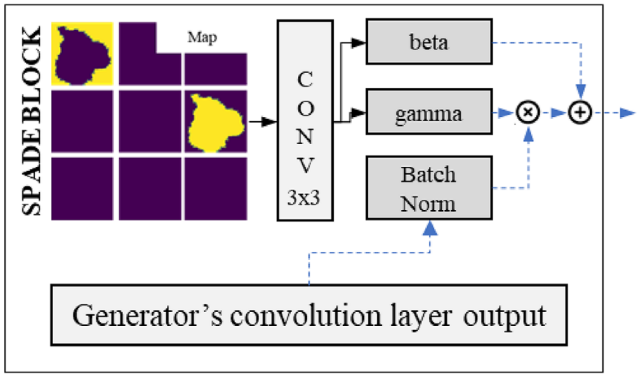

FIGURE 4 The figure shows the role of the SPADE blocks within a Generative Adversarial Network's spatially adaptive normalization. The generated gamma and beta are multiplied and element-wise added to the batch normalization activations. This block enables control over the shape and location of the output activated by the conv layer.

3.4.2 | Implementing the SPADE Block in GauGAN

Here, the focus will be on the implementation and function of the SPADE (Spatially-Adaptive Normalization) block within GauGAN. This section will emphasize how SPADE improves the model's ability to synthesize detailed textures and maintain spatial coherence in generated images. Figure 4 depicts how spatially adaptive normalization, or SPADE, inputs the one-hot-labeled segmentation map, projects it into an embedding space, and then uses convolution to generate gamma and beta tensors. Unlike conditional normalization methods, in which gamma and beta are vectors, in SPADE, they are tensors with a spatial dimension. The generated gamma and beta are added to the normalized activations by element-wise multiplication. Finally, the output parameters of the SPADE block are activated via convolution to efficiently transmit meaningful data to the network.

Traditional normalization techniques like batch normalization or instance normalization standardize feature maps to have a mean of zero and a variance of one. This indiscriminate normalization, however, has the potential to wash away critical spatial details. SPADE overcomes this by conditioning the normalization process on the semantic segmentation maps, ensuring that the normalization is tailored to the image's semantic information.

3.4.3 | Discriminator and generator architecture with SPADE in GauGAN

As depicted in Figure 2, the generator accepts a random vector and a one-hot-labeled segmentation map as inputs and uses them to generate a realistic image. The generator is composed of multiple layers, each of which performs upsampling and normalization operations, as shown in Table 2. For the activation function, we used LeakyReLU, which uses a linear function with a small slope to return the input value as the output when the

TABLE 2 Generator and discriminator layer configuration.

Layer name	Generator		Discriminator	
	Type	Parameter	Type	Parameter
Input	Latent vector	[256]	Images	[224,224,3]
	Mask images	[224,224,10]		[224,224,3]
Convolution	Up sample	[7,7,512] [14,14,512] [28,28,256] [56,56,128] [112,112,64]	Down sample	[112,112,64] [56,56,128] [28,28,256] [14,14,512]
Normalize	Instance normalization			
Activation	LeakyReLU	alpha = 0.2	LeakyReLU	alpha = 0.2
Output	Conv2DTranspose	[224,224,3]	Conv2d	[14,14,1]

input is positive and a linear function with a small slope when the input is negative. Instance normalization is a technique used in CNNs for deep learning models, and in image processing, it is typically applied before the activation function to adjust the distribution of input data. In contrast to the generator, the discriminator is subjected to downsampling, and the same normalization and activation functions applied to the generator are applied to it. The network is built by concatenating the fake and real images generated by the generator and using them as inputs, as well as by updating the losses for the final output.

3.4.4 | Loss function tuning in GauGAN: Optimizing for enhanced image generation

An appropriate loss function's selection is essential for developing a GauGAN model tailored to tomato disease data [35]. Therefore, in this study, we added various loss functions to the loss combination proposed in the paper by Park et al [12]. Also, quantitative metrics such as SSIM (structural similarity index map) and PSNR (peak signal-to-noise ratio) values [36], in addition to qualitative evaluation, were used to assess the optimal combination of loss functions.

The generator and discriminator loss functions of the pix2pix network proposed by Zhu et al. [11] are the principal loss functions used to train GauGAN. The GAN model's objective function is represented by Equation (1). The generator model (G) accepts the source image and the target image as inputs and generates a new image in the style of the target image. Given that the source image is 1 and the fake image is 0, the discriminator is trained to maximize D such that $\log D(x)$ has a probability close to 1. In contrast, the generator is updated to have a probability close to 0 for each fake image it generates, and this process is repeated to update the overall loss function so that the discriminator can predict a generator-generated image as the true image.

$$\min_G \max_D L(G, D) = E_{x \sim pdata(x)} \left[\log D(x)\right] + E_{z \sim pdata(z)} \left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right]. \quad (1)$$

In GauGAN, we replaced the least squared loss function of the discriminator with the hinge loss term, but the performance was higher when the least squared loss was applied. In addition, for GauGAN, we added feature matching loss to generate high-resolution images. Feature matching loss is added to the cost function of a generator to minimize the statistical function between the features of the real image and the features of the generated image. The loss is calculated as the sum of the weighted mean square errors between the feature maps in each layer of the VGG model. The feature matching loss function is constructed as follows:

$$L_{FM} = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot ||D^{(i)}\left(x^{(i)}\right) - D^{(i)}\left(G\left(z^{(i)}\right)\right)||_1, \quad (2)$$

where N denotes the size of the mini-batch, and $D^{(i)}$ denotes the output for layer i of the discriminator. Let $x^{(i)}$ be the output for layer i of the real image and $G(z^{(i)})$ be the output for layer i of the fake image generated by the generator.

To compare the performance of the GauGAN model with those of other models, we also consider the perceptual loss function in this study. Perceptual loss [36] is a loss function designed to supplement MAE (mean absolute error) and MSE (mean squared error). The perceptual loss function directly calculates the perceptual loss as the MSE between feature maps using a user-supplied pre-trained model. The structure of the perceptual loss function is given in Equation (3).

$$L_{PER} = \frac{1}{N} \sum_{i=1}^{N} ||\emptyset\left(x^{(i)}\right) - \emptyset\left(G\left(z^{(i)}\right)\right)||_2^2, \quad (3)$$

where $\emptyset$ denotes the high-level feature representation extracted from the i-th real and fake images using a pre-trained neural network. $||\ ||_2^2$ denotes the squared L2 norm, which measures the Euclidean distance between the feature representations. Finally, the variational encoder, which determines the input data for the generator, applies KL (Kullback–Leibler) divergence to optimize the difference between two probability distributions and to generate a latent vector. The equation for the KL divergence is

$$L_{KL} = \int q(z|x) log\left(\frac{q(z|x)}{p\left(z\right)}\right) dz, \quad (4)$$

where $q(z|x)$ is the Gaussian distribution for the mean and variance computed by the variational encoder, and $p(z)$ is the standard normal distribution. With this expression, the variational encoder is trained to have a normally distributed mean and variance. The final loss function applied in this study is as follows, where λ_1, λ_2, λ_2 are 0.1, 0.1, and 0.5, respectively.

$$L_{gau} = L(G, D) + \lambda_1 L_{FM} \text{ or } \lambda_2 L_{PER} + \lambda_3 L_{KL}. \quad (5)$$

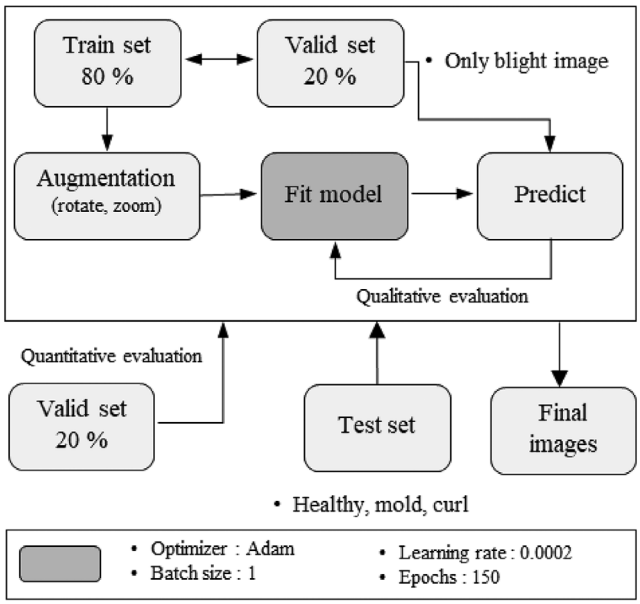

FIGURE 5 The GauGAN network training procedure. The training and validation datasets were split 8:2, and the parameters applied to the training are shown below.

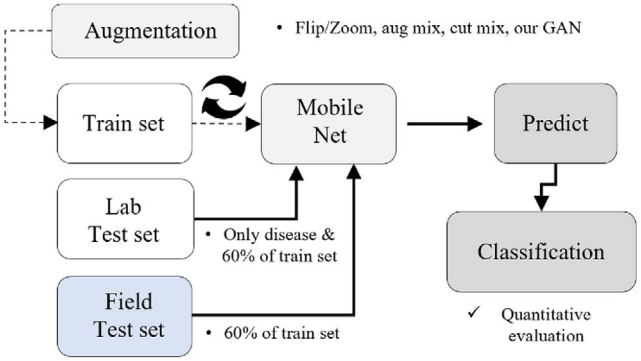

FIGURE 6 Classification performance evaluation process for Lab and field dataset. Split training and validation data 4:6 and utilize mobile net to evaluate classification performance.

4 | RESULTS

4.1 | GauGAN and classification model training detail method

4.1.1 | GauGAN-based image generation

The training and prediction process for the GauGAN network is depicted in Figure 5. The data for training and validation were divided 8:2, and the training data were augmented with rotation and magnification capabilities. During training, visual evaluation was used every 5 epochs to determine quality of the images produced by the validation data. All models were trained using an Adam optimizer with a fixed learning rate of 0.0002, a batch size of 1, and a training duration of 150 epochs. The disease images were trained only on images corresponding to early blight, and the entire training process required approximately six hours on an NVIDIA GeForce RTX 2060 SUPER. The programming and implementation were performed in Python using TensorFlow libraries within a Jupyter Notebook framework. The augmented data were used to generate augmented images with a mask corresponding to the classification, excluding early blight.

4.1.2 | Disease classification for generated images

To evaluate the effect of data augmentation and the classification performance of the generated images, we established the classification procedure depicted in Figure 6. MobileNet, a lightweight model suitable for mobile applications and real-time detection, was used to perform image classification [37]. In the evaluation phase, the hyper-parameters were tuned using only the test data for early blight to assess the prediction accuracy of the blight. In this approach, 200 images of the early blight were selected from the training data, creating an imbalanced data structure, and the remaining 200 images were used for prediction and assessment. For further validation of this approach, an imbalanced data structure was created for each disease within the plant village dataset. The training and validation data were then divided at a ratio of 4:6, and the model's accuracy was evaluated. Similarly, the field data was analyzed in the same process.

4.2 | Results based on lab and field datasets

4.2.1 | Generated images using GauGAN based on lab dataset

Figure 7 depicts the results generated by the GauGAN model under each condition. The base GauGAN is the image generated by the model implemented using the Keras tutorial code, and loss tuning is the result of applying the loss function of pix2pix's generator and discriminator. In the qualitative evaluation, the images did not differ significantly, but when loss tuning was applied, the learning was more stable. The feature matching model implements an additional feature matching loss, whereas the modified model implements perceptual loss. With the training data, the modified model's prediction result did not differ significantly from the control; with the validation data, it predicted the disease area more accurately than the control.

4.2.2 | Comparative evaluation results for generated images

Table 3 shows the results from the quality assessment of images generated by the GauGAN model and the classification performance in the quantitative evaluation of the GauGAN model. We used and SSIM values to quantitatively evaluate the quality of the generated images. PSNR measures the amount of

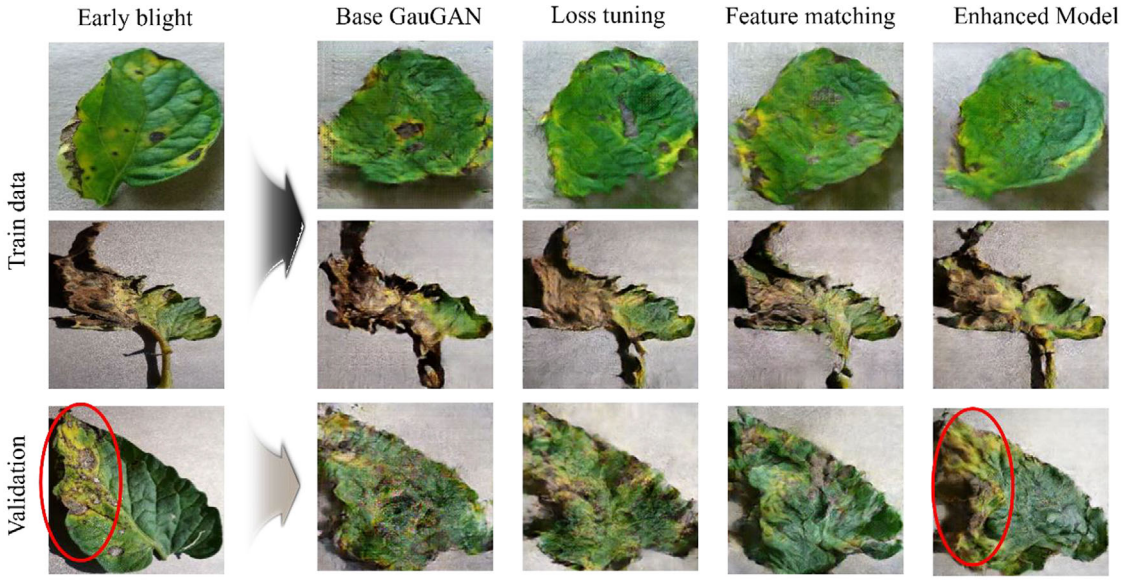

FIGURE 7 A comparison of the images generated under each condition. Compared to the pix2pix algorithm, the generated images are more diverse, and it is possible to distinguish between disease and leaf areas. The red coloring indicates that the diseased areas are well simulated.

TABLE 3 Performance metrics for various model configurations.

Metric	Base GauGAN	Loss tuning	Feature matching	Modified model
SSIM	0.3412	0.3815	0.4025	0.4214
PSNR(dB)	17.58	18.30	18.67	19.14
F1-score	0.8472	0.8826	0.9528	0.9949
Accuracy	0.7350	0.7900	0.9100	0.9900

TABLE 4 Performance metrics for imbalanced data.

Accuracy	Val 1 (Early blight)	Val 2 (Leaf mold)	Val 3 (Leaf spot)
Original images	0.7500	0.6800	0.6500
GauGAN images	0.9000	0.7900	0.8000

quality loss in an image; high values for both variables indicate high image quality. SSIM measures the correlation between two images in terms of luminance, contrast, and structure. The modified model had high values for both metrics. The robustness results of the modified model are also shown in Table 4. For all the validation datasets that generated imbalances for each disease, we found that data augmentation improved performance over the original dataset. Finally, the GauGAN model was validated, and the modified model for the tomato disease data was chosen.

To assess the efficacy of data augmentation on images generated by the modified model, we evaluated the classification performance using conventional techniques (Flip and Zoom) and more advanced techniques (Mix-up and Cut-mix). The results from the Mix-up and Cut-mix techniques are depicted in Figure 8.

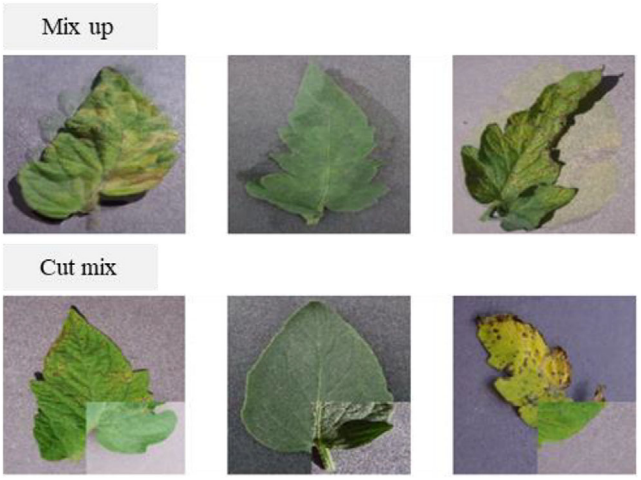

FIGURE 8 Example of the result of applying Mix-up and Cut-mix. Shows the results of augmentation applied to training data when training a classification model.

Table 5 shows the classification performance as evaluated by the F1-score and accuracy metric. The data augmentation

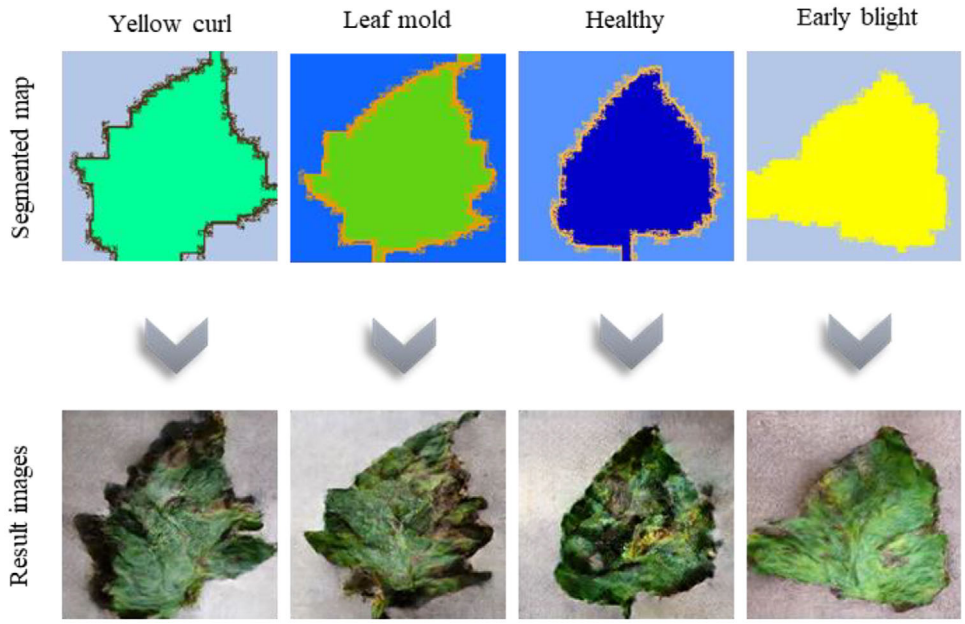

FIGURE 9 Resulting images predicted by the modified model. The left figure shows a labeled segmented map, and the right figure shows that the images are generated while preserving the geometry.

TABLE 5 Comparing results across different augmentation techniques.

Metric	Base	Flip/Zoom	Mix-up	Cut-mix	Modified model
F1-score	0.8603	0.8338	0.9010	0.9473	0.9949
Recall	0.7550	0.7150	0.8200	0.9000	0.9900
Accuracy	0.7550	0.7150	0.8200	0.9000	0.9900

images generated by the modified model were used to train the classification model. The effects of the various traditional data augmentation techniques on classification performance were evaluated and are shown in Table 5. When the modified model, which combined optimized parameters and augmentations, was compared with the base model without any data augmentation, an improvement of approximately 20% was observed. The modified model also outperformed the other approaches tested, achieving an F1 score of 0.994 and an accuracy of 0.99.

The reason for the equivalence of recall and accuracy scores in this experiment is solely the focus on evaluating early blight disease. Consequently, the results in Table 5 represent the assessment of 200 early blight images that were completely independent of the training process. These results support the efficacy of the data augmentation techniques proposed in this study in improving the robustness and performance of classification models.

Figure 9 depicts segmentation map labeled with color assigned to each disease category, as well as the resultsgenerated by the modified model. In the segmentation map, the color differences are easily discernible. In an image generated by the GauGAN model, the leaf shape on the segmentation map is reflected as accurately as possible, resulting in a realistic and structurally consistent leaf image. The overfitting issue in the validation tests mentioned in Section 3 was resolved by tuning the losses of the GauGAN model.

4.2.3 | Validation results for field datasets

To further validate the model presented in this study, we evaluated its applicability to field data. Data collected in the field differ significantly from the PlantVillage data used to develop the model, including in geometric diversity, complexity, and irregular backgrounds. In fact, leaf area, not background, is the most important factor for disease detection. To improve the generalization performance with the field data, a Grabcut module was applied to remove unnecessary background before training. In this study, focus was placed on leaf diseases. Figure 10 shows an input image with this removed background and the modified image generated by the modified model. The qualitative evaluation carried out through visual inspection shows that images with different colors were generated but the shape of the disease region was maintained.

The classification performance was also evaluated, and we found that the images generated by the modified model performed well, as evidenced by the results shown in Table 6. Recall is a measure that focuses specifically on a model's performance in identifying the disease class. Compared with other data augmentation techniques, the recall score improved by about 10% when using the modified model, further validating the technique proposed in this study. Additionally, recent augmentation techniques, specifically the RandAugment method [16], have been further evaluated for field data. The use of GauGAN for data augmentation demonstrated more than a 4% higher result compared to these methods.

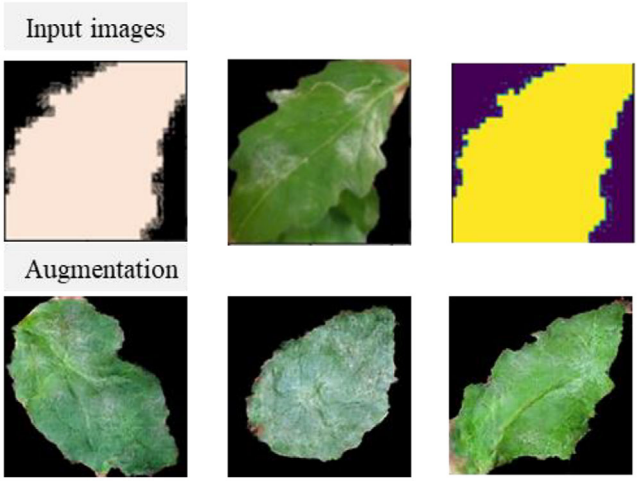

FIGURE 10 Input images of field data and images generated by GauGAN. For input images, from left to right, segmented map, real image, and one-hat labeled map. The augmented results below show some randomly selected images.

5 | CONCLUSIONS

In this study, we investigated augmentation techniques for plant disease data using GANs, specifically the GauGAN algorithm. Our motivation in this research was to solve the problem of data imbalance that degrades the generalization performance of deep-learning classification models. The GauGAN model showed high performance in generating disease images while maintaining the background based on a mask image. Furthermore, images augmented by the GauGAN model improved the performance of MobileNet-based classification models.

The GauGAN model has a complex structure that requires a segmented mask image and a one-hot encoded segmentation map with label information. This complexity can lead to several challenges in training the model. In this study, we found that overfitting occurred when the latent vector generated by the variational encoder was updated in the prediction phase. In addition, overfitting occurred when classification of the input labels was unclear or the label information was not entered properly. To solve those problems, we established a preprocessing step for the input images. With an effective preprocessing technique, we were able to solve the overfitting problem and implement a data augmentation model with an appropriate combination of GAN loss and perceptual loss. In particular, images augmented with this configuration improved the F1 score by more than 10% compared with the conventional method. The performance with field data was similarly improved.

However,the models proposed in this study are not generalizable to all datasets. Our research focused on applying GAN-based data augmentation techniques to botanical datasets and updating the GauGAN algorithm to suit them specifically. Additional tuning and validation would be required to extend this technique to datasets in other domains.

TABLE 6 Results of comparing augmentation techniques on field data.

	Base	Flip/ Zoom	Mix up	Cut mix	RandAugment	Ours
F1-score	0.9234	0.9452	0.8754	0.8042	0.9077	0.9412
Recall	0.7175	0.7651	0.5885	0.5245	0.6994	0.8125
Accuracy	0.9354	0.9321	0.9051	0.8621	0.9144	0.9515

In conclusion, this study not only validated the utility of Gaussian-based data enhancement in an agricultural setting, but also hints at its broader applicability in fields such as healthcare and environmental monitoring. In the future, we aim to fine-tune these techniques and integrate them with other data enrichment strategies to create more robust and generalizable models.

AUTHOR CONTRIBUTIONS

Seung-Beom Cho: Conceptualization; data curation; formal analysis; methodology; software; validation; writing—original draft; writing—review & editing. **Yu Cheng**: Methodology; writing—review & editing. **Sanghun Sul**: Conceptualization; methodology; supervision.

ACKNOWLEDGEMENTS

This work was supported by a Korean Institute of Design Promotion (KIDP) grant funded by the Korean government (MOE, MOTIE) (Design and Emerging Technology Integrated Education Program for Cultivating Innovative Talents).

CONFLICTS OF INTEREST STATEMENT

No potential conflict of interest relevant to this article was reported.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

ORCID

Seung-Beom Cho https://orcid.org/0000-0002-7431-8789

REFERENCES

1. Paul, S.G., et al.: A real-time application-based convolutional neural network approach for tomato leaf disease classification. Array 19, 100313 (2023)
2. Talasila, S., Rawal, K., Sethi, G.: Conventional data augmentation techniques for plant disease detection and classification systems. In: Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021, pp. 279–287. Springer Nature, Singapore (2022)
3. Nagaraju, M., Chawla, P., Kumar, N.: Performance improvement of deep learning models using image augmentation techniques. Multimed. Tools Appl. 81(7), 9177–9200 (2022)
4. Goodfellow, I., et al.: Generative Adversarial Nets. In: Advances in Neural Information Processing Systems, vol. 27. MIT Press, Cambridge, MA (2014)
5. Jia, H., et al.: BlazeStyleGAN: A real-time on-device StyleGAN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4689–4693. IEEE, Piscataway (2023)

6. Lyu, Y., et al.: ARU-GAN: U-shaped GAN based on attention and residual connection for super-resolution reconstruction. Comput. Biol. Med. 164, 107316 (2023)
7. Thangaraj, R., et al.: Artificial intelligence in tomato leaf disease detection: A comprehensive review and discussion. J. Plant Dis. Prot. 129(3), 469–488 (2022)
8. Min, B., et al.: Data augmentation method for plant leaf disease recognition. Appl. Sci. 13(3), 1465 (2023)
9. Chen, Y., Pan, J., Wu, Q.: Apple leaf disease identification via improved CycleGAN and convolutional neural network. Soft Comput. 27, 9773–9786 (2023)
10. Cui, X., Ying, Y., Chen, Z.: CycleGAN based confusion model for cross-species plant disease image migration. J. Intell. Fuzzy Syst. 41(6), 6685–6696 (2021)
11. Cho, S.-B., et al.: Heterogeneous domain adaptation method for tomato leaf disease classification base on CycleGAN. J. Intell. Fuzzy Syst. 45(5), 8859–8870 (2023)
12. Park, T., et al.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2337–2346. IEEE, Piscataway (2019)
13. Hendrycks, D., et al.: Augmix: a simple data processing method to improve robustness and uncertainty. arXiv preprint, arXiv:1912.02781 (2019)
14. Shen, Y., et al.: High-throughput phenotyping of individual plant height in an oilseed rape population based on Mask-RCNN and UAV images. Precis. Agric. (2023)
15. Lim, S., et al.: Fast Autoaugment. In: Advances in Neural Information Processing Systems, vol. 32. MIT Press, Cambridge, MA (2019)
16. Lee, G., et al.: Enhancing detection performance for robotic harvesting systems through RandAugment. Eng. Appl. Artif. Intell. 123, 106445 (2023)
17. Yun, S., et al.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6023–6032. IEEE, Piscataway (2019)
18. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419 (2016)
19. Fuentes, A., et al.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 17(9), 2022 (2017)
20. Su, D., et al.: Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. Comput. Electron. Agric. 190, 106418 (2021)
21. Enkvetchakul, P., Surinta, O.: Effective data augmentation and training techniques for improving deep learning in plant leaf disease recognition. Appl. Sci. Eng. Prog. 15(3), 3810–3810 (2022)
22. Lamba, S., Baliyan, A., Kukreja, V.: GAN based image augmentation for increased CNN performance in Paddy leaf disease classification. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 2054–2059. IEEE, Piscataway (2022)
23. Cap, Q.H., et al.: Leafgan: An effective data augmentation method for practical plant disease diagnosis. IEEE Trans. Autom. Sci. Eng. 19(2), 1258–1267 (2020)
24. Guo, H., et al.: Sample expansion and classification model of maize leaf diseases based on the self-attention CycleGAN. Sustainability 15(18), 13420 (2023)
25. Zhang, Z., et al.: A high-quality rice leaf disease image data augmentation method based on a dual GAN. IEEE Access 11, 21176–21191 (2023)
26. Min, B., et al.: Data augmentation method for plant leaf disease recognition. Appl. Sci. 13(3), 1465 (2023)
27. Sun, K., et al.: TsrNet: A two-stage unsupervised approach for clothing region-specific textures style transfer. J. Visual Commun. Image Represent. 91, 103778 (2023)
28. Zhang, Z., et al.: A high-quality rice leaf disease image data augmentation method based on a dual GAN. IEEE Access 11, 21176–21191 (2023)
29. Strelcenia, E., Prakoonwit, S.: A new GAN-based data augmentation method for handling class imbalance in credit card fraud detection. In: 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 627–634. IEEE, Piscataway (2023)
30. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, arXiv:1511.06434 (2015)
31. Wu, Q., Chen, Y., Meng, J.: DCGAN-based data augmentation for tomato leaf disease identification. IEEE Access 8, 98716–98728 (2020)
32. Wang, T.-C., et al.: High-resolution image synthesis and semantic manipulation with conditional Gans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8798–8807. IEEE, Piscataway (2018)
33. Hughes, D., et al.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint, arXiv:1511.08060 (2015)
34. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc., C: Appl. Stat. 28(1), 100–108 (1979)
35. Salimans, T., et al.: Improved Techniques for Training GANs. In: Advances in Neural Information Processing Systems, vol. 29. MIT Press, Cambridge, MA (2016)
36. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV2016: 14th European Conference, Proceedings, Part II 14, pp. 694–711. Springer International Publishing, Berlin (2016)
37. Howard, A.G., et al.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint, arXiv:1704.04861 (2017)

How to cite this article: Cho, S.-B., Cheng, Y., Sul, S.: Enhanced classification performance through GauGAN-based data augmentation for tomato leaf images. IET Image Process. 1–11 (2024). https://doi.org/10.1049/ipr2.13069