applied sciences

MDPI

Article

A Novel Generative Adversarial Network Model Based on GC-MS Analysis for the Classification of Taif Rose

Hala M. Abdelmigid [1,*], Mohammed Baz [2], Mohammed A. AlZain [3], Jehad F. Al-Amri [3], Hatim G. Zaini [2], Maissa M. Morsi [4], Matokah Abualnaja [5] and Nawal Abdallah Alhuthal [6]

[1] Department of Biotechnology, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[2] Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[3] Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[4] Department of Biology, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[5] Department of Chemistry, Faculty of Applied Science, Umm Al-Qura University, Makkah 24230, Saudi Arabia
[6] Department of Chemistry, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
* Correspondence: h.majed@tu.edu.sa; Tel.: +966-551-785-254

Abstract: Rose oil production is believed to be dependent on only a few genotypes of the famous rose *Rosa damascena*. The aim of this study was to develop a novel GC-MS fingerprint based on the need to expand the genetic resources of oil-bearing rose for industrial cultivation in the Taif region (Saudi Arabia). Gas chromatography-mass spectrometry (GC-MS) is a widely used analytical technique for determining the volatile composition of distilled rose oil from flower data. Because biosample availability, prohibitive costs, and ethical concerns limit observations in agricultural research, we aimed to enhance the quality of analysis by combining real observations with samples generated in silico. This study proposes a novel artificial intelligence model based on generative adversarial neural networks (GANs) to classify Taif rose cultivars using raw GC-MS data. We employed a variant of the GAN known as conditional stacked GANs (cSGANs) to predict Taif rose's oil content and other latent characteristics without the need to conduct laboratory tests. A hierarchical stack of conditional GANs is used in this algorithm to generate images. A cluster model was developed based on the dataset provided, to quantify the diversity that should be implemented in the proposed model. The networks were trained using the cross-entropy and minimax loss functions. The accuracy of the proposed model was assessed by measuring losses as a function of the number of epochs. The results prove the ability of the proposed model to perfectly generate new real samples of different classes based on the GC-MS fingerprint.

Keywords: conditional generative adversarial network; Taif rose; gas chromatography-mass spectrophotometry; StackGAN

Citation: Abdelmigid, H.M.; Baz, M.; AlZain, M.A.; Al-Amri, J.F.; Zaini, H.G.; Morsi, M.M.; Abualnaja, M.; Alhuthal, N.A. A Novel Generative Adversarial Network Model Based on GC-MS Analysis for the Classification of Taif Rose. *Appl. Sci.* **2023**, *13*, 3052. https://doi.org/10.3390/app13053052

Academic Editor: Zhengjun Qiu

Received: 12 December 2022
Revised: 6 February 2023
Accepted: 20 February 2023
Published: 27 February 2023

1. Introduction

Rosa x damascena Herrm is one of the most famous representatives of the genus *Rosa* L. It is a hybrid rose derived from *Rosa gallica* and *Rosa moschata*. Additionally, it is known as the Damask rose, the Bulgarian rose, the Turkish rose, the Taif rose, the Arab rose, and the Castile rose [1]. It has been cultivated in 18,000 different varieties [2,3]. Because of its high ornamental value and valuable essential oil, the Damask rose is widely cultivated in many countries. The essential oil of *R. damascena* has been used in perfumery, cosmetics, aromatherapy, and various medicinal applications since ancient times. Some authors believe that rose oil production depends only on a few genotypes of the famous rose *R. damascena*. According to the current global market for rose oil, Bulgaria, Turkey, Iran, and India are leading suppliers, alongside Saudi Arabia (SA). The Taif region of SA is a major producer and relies on only one genotype. There has been increasing concern regarding

the need to increase the genetic resources of oil-bearing roses for industrial cultivation [4]. Saudi Arabia has developed *R. damascena* exclusively through clonal selection, thereby preserving its traditional odor and composition by avoiding crossbreeding. Over the past decade, molecular breeding has proven to be a highly effective method for improving oil roses [5]. According to international standards [6] governing the composition of oil roses, when hybridizing *R. damascena* between or within varieties, desired traits must be introduced without altering the volatiles of the flowers. Therefore, it is crucial that a high-throughput method be used to determine the relative composition of rose oils collected from various breeding lines, as well as accessions of segregated and natural rose populations.

Gas chromatography–mass spectrometry (GC-MS)—an instrumental technique used to determine the components of essential oils—consists of a gas chromatograph (GC) combined with a mass spectrometer (MS) to separate, identify, and quantify complex mixtures of chemicals. In the field of quantitative GC-MS, which is widely used to analyze floral volatiles, flower data can be extrapolated to the volatile composition of distilled rose oil based on specific ion patterns. Furthermore, GC-MS spectral data can be analyzed to create fingerprint templates, and a fingerprint image generation program can be used to structure the complex instrumental data. Thus, it is ideal to analyze hundreds of compounds with relatively low molecular weight [7]. For complex samples, peak overlap is likely, due to the background, baseline offset, and some overlapped or embedded peaks, even under ideal experimental conditions. Often, these problems lead to incorrect similarity matches in the MS library, resulting in incorrect component identification. Consequently, the primary objective of research in this area is to determine the constituents of the compounds and resolve them.

Machine learning algorithms have been used to create kinetic and genome-scale models for data-driven predictions and conclusions [8–10]. Existing models are used to determine the features that are essential in a network using machine learning algorithms [9]. It is also possible to map metabolic models onto genomic, proteomic, and metabolomic data using machine learning techniques, such as clustering [11] or support-vector machines [12]. Using multiview machine learning algorithms, different omics data and metabolic model data can be combined, which may include separate machine learning algorithms that analyze each omics layer and aggregate them, utilizing methods developed through multilayer network theory [13].

Rather than using human-engineered features, deep learning models use deep neural networks and complex architectures to extract information directly from the raw features. Deep learning is widely accepted as one of the most rapidly developing methodologies in the scientific field [14]. Convolutional neural networks (CNNs) are widely applied in image classification, ranging from natural image classification to computer-aided diagnosis [15,16], but the generalization of CNNs is limited when applied to validation or testing datasets because they require large training datasets. In practice, in silico generation has been successfully used in computer vision for data enhancement even when the number of biological samples is limited. The use of samples generated in silico has been combined with observations made from real samples to artificially increase the number of observations. In this study, we augmented real GC-MS data with newly generated samples, whose distribution mimics the original data distribution in their original metabolite space. Data modeling traditionally relies on deep-learning-based generative adversarial networks (GANs) for the generation of photorealistic images, rather than the priors underlying such data.

GANs were introduced as deep learning tools by Goodfellow et al. in 2014 [17]. A GAN is a generative model that uses deep neural networks in an adversarial setting. Specifically, a GAN uses adversarial methods to learn generative models of the data distribution. This has become one of the hottest research areas in artificial intelligence, as one of the most successful generative models in recent years. Since its initial proposal, the GAN has attracted considerable attention owing to its exceptional performance. In addition to its excellent performance as a generative model, the GAN integrates deeply into all facets of

deep learning, opening up an array of new research directions and applications [18,19]. Generative adversarial networks, as a technique for augmenting data scarcity, provide the ability to simulate existing images, so they are particularly promising for overcoming data scarcity [20,21]. By constructing an adversarial network, the GAN trains a generator and a discriminator. Conditional GANs (cGANs) are deep learning neural networks with additional parameters that depend on the project requirements. In addition, labels are used in the discriminator inputs so that the discriminator can correctly classify the input and cannot be easily filled in by the generator [22].

In this study, a novel artificial intelligence model was developed to classify Taif rose images based on their GC characteristics. Using this model, the oil and other latent characteristics of roses can be predicted without requiring laboratory testing. However, achieving this goal through traditional ANN models is difficult because of the relatively small dataset available for the Taif rose images and their GC analysis. Therefore, this study employs a state-of-the-art generative approach, namely, a conditional stacked GAN model (cSGAN). StackGANs are generative adversarial networks (GAN variants) that use hierarchical stacks of conditional GANs to generate images. As the name implies, StackGANs have two stacked GANs for creating a network that can generate high-resolution images. Using this approach, we evaluated 10 rose cultivars from different locations within the Taif region of Saudi Arabia.

To the best of our knowledge, this is the first study to propose the use of a GAN to learn the chemical component patterns of Taif rose flowers directly from the raw GC-MS data. The key contributions of this study are as follows:

1. Development of a novel GC-MS fingerprint based on the need to expand the genetic resources of oil-bearing roses for industrial cultivation in the Taif region (Saudi Arabia).
2. Providing agricultural researchers with a means of overcoming the problem of the shortage of observations caused by limited access to biosampling techniques, prohibitive costs, or ethical concerns. A more accurate data analysis may be possible by integrating samples generated in silico with real observations.
3. Evaluating cSGANs as a method for generating a variety of realistic classes by analyzing the GC–MS fingerprints of rose oils derived from populations of cultivated *R. damascena* capable of producing oils.
4. Analysis of the dataset provided for the development of a cluster model that quantifies the diversity that needs to be incorporated into the proposed model.

The remainder of this paper is organized as follows: Section 2 describes related work, Section 3 presents the materials and techniques used for obtaining data, and Section 4 describes the algorithm and methodologies used in the study. In Section 5, the results obtained using the previous techniques are presented. An assessment of the model is provided in Section 6, along with a discussion of the results. Finally, the conclusions of this study are presented in Section 7.

2. Related Work

Research on GANs has primarily focused on unsupervised settings that utilize unlabeled images. Conditional GANs [23] are able to generate images that are based on a set of attributes. According to Schlegl et al. [23], attributes may be fed to both the generator and the discriminator such that the generator can generate images based on these attributes. An alternative approach, proposed by Odena et al. [21], utilizes auxiliary classifier GANs (AC-GANs), in which the discriminator reconstructs side information as part of the classification. This line of research focuses on supervised settings, which assume that all images possess attribute tags, regardless of the approach employed.

GANs can render deep representations of real data by implementing a minimax game process [17]. This technology can achieve state-of-the-art performance in various applications, such as image synthesis, super-resolution, visual sequence prediction, and semantic image painting. Thus, this emerging method has received increasing attention for the detection of outliers. Using the same generative model, Schlegl et al. [23] investigated

the posterior probabilities of the abnormal markers in medical images. Zenati et al. [24] employed a network structure based on BiGAN [25] to jointly train the mapping from an image to latent space. In contrast, Akcay et al. [26] presented a subnetwork containing encoders, decoders, and encoders to reduce the computational complexity associated with remapping latent vectors. Deecke et al. [27] initialized multiple latent vectors from various locations in the latent space to solve the nonconvexity of the underlying optimization problem. Nonetheless, they defined a GAN as a feature extractor or reconstructor, which differs considerably from our approach. Additionally, we extend the GAN-based models beyond a single generator model (SO-GAAL) to a model with multiple generators with different objectives (MO-GAAL) to address the mode collapse issue.

Several studies have reported that AC-GANs are effective in classifying images. Schlegl et al. [23] used AC-GANs to enhance a CNN liver lesion classifier using three classes. It was found that a discriminator independent of an AC-GAN can be used to classify images, resulting in a 2% reduction in classification accuracy. In contrast, their CNN classifier was designed to accurately categorize the livers, whereas their AC-GAN architecture was the same as that used by Soleimani et al. [28]. However, the literature has demonstrated promising results from all-in-one architectures that combine both generation and classification [9,29,30]. An AC-GAN variant has been proposed by Yang et al. [29] for hyperspectral imaging, plant segmentation, and image classification. Furthermore, the Wasserstein GAN has been applied with gradient clipping to classify a specific signal [31]. We have been motivated to determine how AC-GANs perform under controlled conditions with similar hyperparameters in the same domain, as compared to standard CNNs. As a result of exploring methods to overcome the limited size of datasets, AC-GANs were further adapted for image classification.

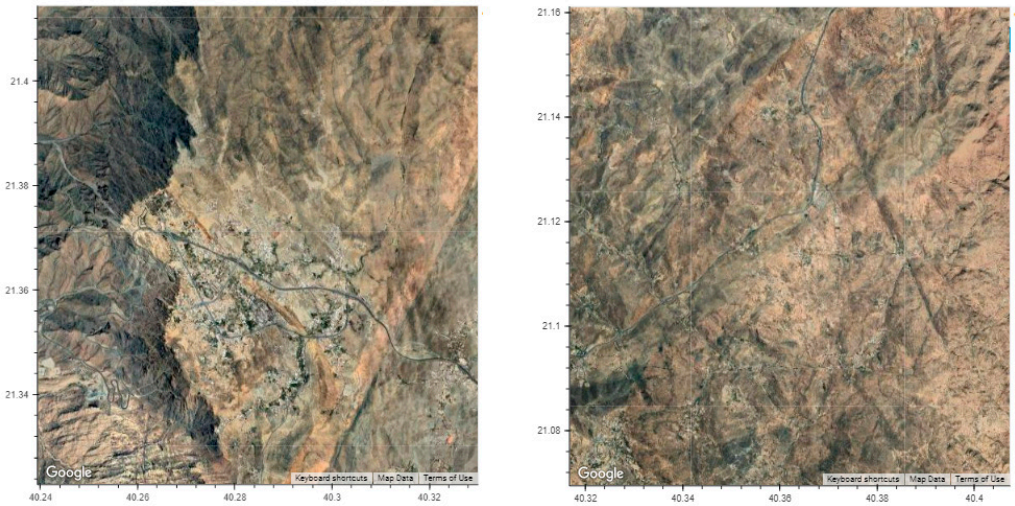
With the recent advancements in agriculture and plant science, artificial images based on GANs have become increasingly popular. GANs are primarily concerned with translating between two domains (A and B), where a domain refers to a collection of samples, such as images, the distribution of which is implicitly determined by the GAN. Several GAN approaches in the literature differ in the selection of domains A and B in the training dataset, with images generated when applying the training method from domain B [32]. Conditional GANs (cGANs) can be used in agricultural applications to discover powerful generators based on a set of inputs and outputs [33]. There is evidence that these networks have achieved good results in various applications in the domain adaptation field [18,22,34]. However, they are rarely used in plant science. Plant data have already been successfully used in cGANs to improve data for a few specific applications. To achieve this goal, Zhu et al. [20] synthesized new real-looking images of plants using segmentation masks on the input side of the cGAN. To acquire high-resolution images, Zhang et al. developed a stacked generative adversarial network (SGAN) architecture [35] using GAN models. It was difficult to generate high-resolution images from text descriptions, owing to the complexity of the learning process. This difficulty was often caused by a wide range of variations in the correspondence between natural languages and images. The StackGAN algorithm can create high-resolution images of 256 $\times$ 256 pixels; however, it has been plagued by problems such as unintelligible images and mode collapse because of unstable learning. A conditional generation task becomes more problematic as the image resolution increases.

To the best of our knowledge, no previous study has been conducted on the utility of agricultural data pairs in cSGANs, where the domains differ in the GC-MS data (e.g., retention time, molecular weight, and peak percentage), as in this study. Nevertheless, our research is methodologically related to the analysis of facial aging, which involves predicting faces several years into the future using cGANs [36]. Similarly, cGANs can help maintain certain traits in plants, just as they are crucial for maintaining the characteristics that determine an individual's identity.

3. Materials and Methods

3.1. Study Area

This study was conducted in Taif, Makkah Province, Saudi Arabia, on the western slopes of the Al-Sarawat Mountains, at an elevation of 1879 m (6165 feet). In terms of latitude and longitude, Taif is located at 21°262′03″ N and 40°212′03″ E, respectively. The Taif region reaches a maximum temperature of 34 °C in summer and a minimum temperature of 10 °C in winter. Approximately 25,500 hectares of agricultural land is located in Taif, with rainfall occurring almost continuously throughout the year, with a higher percentage in spring and late autumn. These farms are primarily located on the Al-Shafa and Al-Hada plateaus (Figure 1).



(**a**) (**b**)

Figure 1. Geographical locations of the 10 Taif rose farms (Al-Hada (**a**) and Al-Shafa (**b**)) from which the samples were collected.

3.2. Sampling

During the 2021 rose harvest season (early morning hours), rose flowers were collected from 10 oil-bearing rose accessions as part of a traditional rose flower collection practice. The flowers were immediately frozen in liquid nitrogen and stored at −80 °C for further analysis.

3.3. GC-MS Analysis

Approximately 20 mg of rose essential oil was extracted from 100 g of the flower sample and dissolved in 1 mL of N-hexane for four hours. For each sample, three replicates were used. As described by Xiao et al. [37], essential oil samples were dissolved in dichloromethane for GC-MS analysis. GC-MS analysis was performed using a Thermo Scientific Trace GC Ultra/ISQ Single-Quadrupole MS with a TG-5MS fused-silica capillary column (30 m, 0.251 mm, 0.1 mm film thickness) Markham, ON, CA. Owing to the simplicity and rapidity of this method, it is possible to accurately identify the origin of rose varieties. Table 1 lists the parameters used in this analysis.

After headspace injection, the volatile organic components in the rose samples were detected quickly by heating them directly. By analyzing the gas-phase ion migration spectra and fingerprint chromatograms, the instrument can determine the differences between the volatile organic compounds in each sample. Based on the percentage of the relative peak area, all identified components were quantified. The compounds were tentatively identified based on a comparison of their relative retention times and mass spectra with NIST and Wiley library data obtained from the GC-MS system [38]. An example of GC-MS analysis that we received from the laboratory is provided in Figure 2 and Table 2.

Table 1. Various parameters of GC–MS analysis used in this research.

Parameter Settings for Thermo Scientific, Trace GC Ultra/ISQ Single-Quadrupole MS	
Analysis time	60 min.
Column type	TG-5MS fused-silica capillary column
Ionization energy	70 eV
Carrier drift gas	Helium (He)
Carrier gas flow rate	1 mL/min.
Initial MS temperature 40 °C/3 min	40 °C
Increasing rate of temperature	5 °C/min (hold 5 min)
The injector and MS final temperature	280 °C

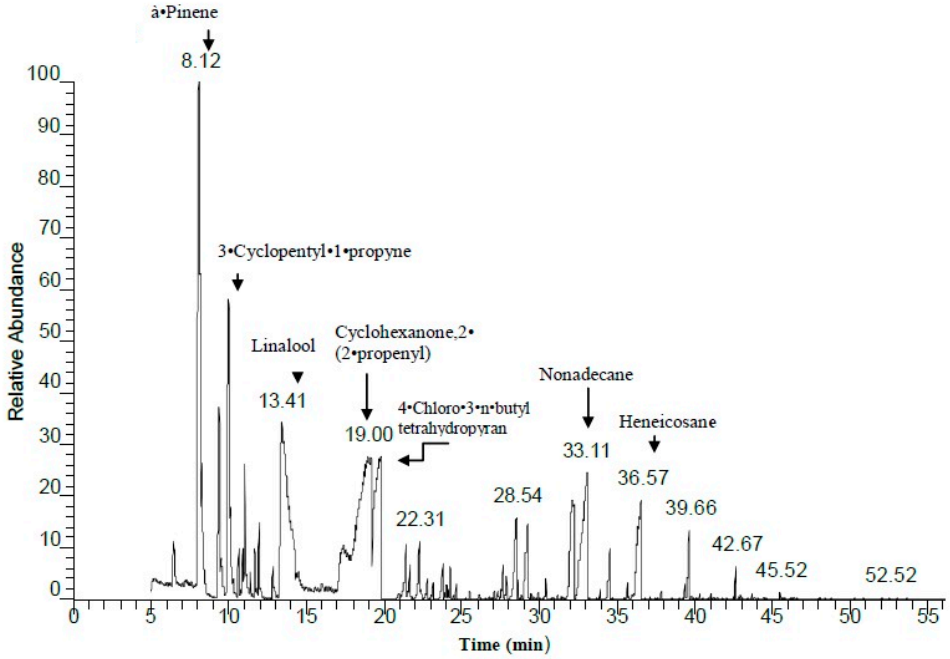


Figure 2. Representative sample of the GC-MS fingerprints generated by GC-MS analysis of the studied rose oils.

Table 2. Example of GC–MS data of rose oil composition. RT = retention time, MW = molecular weight, MF = molecular formula. Area% = the percentage of the component.

R_T	Mw	MF	Area%	Compound Name
5.16	232	C14H16O 3	0.83	Ethyl- 2-[(benzyloxy)methyl]cycloprop-2-ene-1-carboxylate
6.46	210	C15H30	1.05	2,4,6,8-Tetramethyl-1-undecene
8.11	136	C10H16	18.43	á-Pinene
8.29	136	$C_9H_{12}O$	1.06	Spiro[cyclopropane-1,6′[3]-oxatricyclo [3.2.1.0(2,4)]octane]
9.41	120	C 9H12	4.09	1,8-Nonadiyne (CAS)
9.56	136	C10H16	0.48	Sabinene
10.02	108	C8H12	8.42	3-Cyclopentyl-1-propyne -6
10.13	100	$C_6H_{12}O$	0.23	Cyclopentanemethanol
10.68	136	C10H16	0.64	çTerpinene
10.95	134	C10H14	0.55	Benzene,1-ethyl-2,4-dimethyl (CAS)
11.07	136	C10H16	1.65	á-Phellandrene

Table 2. *Cont.*

R_T	Mw	MF	Area%	Compound Name
11.40	136	C10H16	0.26	Bicyclo-[3.1.1]-hept-2-ene,3,6,6-trimethyl (CAS)
11.71	136	C10H16	0.61	1,3,6-Octatriene,3,7-dimethyl-,(E)-(CAS)
12.00	136	C10H16	0.90	ç-Terpinene
12.88	136	C10H16	0.90	ç-Terpinene
13.41	154	C10H18O	9.88	Linalool
14.26	218	$C_9H_{15}BrO$	0.73	1á-Bromo-3Aà,4à,5,6,7,7Aà-Hexahydroindan-4-ol
17.30	154	C9H14O2	1.40	MethylBicyclo [3.1.0]hexane-6-acetate

4. The Proposed Model

Several annotated natural image datasets are available in the fields of deep learning and computer vision for use with classifier models. These datasets can be used to identify roses based on their gas-phase ion migration spectra. It is evident that developing large-scale professional image datasets and integrating numerous annotations would undoubtedly require a significant amount of manpower and materials, owing to privacy restrictions, industry standards, and unintegrated information systems. Based on a small set of marked and unlabeled data, semi-supervised classification generates a set of data category labels by using a GAN discriminator network.

4.1. Description of the Model

As a deep learning framework, generative adversarial networks (GANs) involve training two models simultaneously: a generative model G, and a discriminative model D. G aims to capture the distribution of some target data (e.g., distributions of pixel intensity in images). D aids G by examining the data generated by G and comparing the generated data with the "real" data. According to Goodfellow et al. [17], a GAN is a pair of simple neural networks; however, in practice, it can be any pair of generative–discriminative networks [39].

In the original works [27,39], the GAN is analogized as a counterfeit money production process, where $\mathfrak{S}$ plays the role of a counterfeiter in training, whereas $\mathfrak{D}$ (the bank) tries to identify fake bills, and in doing so helps $\mathfrak{S}$ develop its skills. Specifically, let $x \sim p_{data}$ be featured by defining actual bills, while $\mathfrak{S}(z)$ features $\mathfrak{S}$ created from some noise distribution $z \sim p_z$. In addition, let $\mathcal{J}$ be a quantitative metric that measures how real a bill is. Then, $\mathfrak{D}$'s role is to lower $\mathcal{J}(\mathfrak{S}(z))$ (the score of a fake bill) while increasing $\mathcal{J}(x)$ (the score of a real bill) for more accurate identification. However, the goal of $\mathfrak{S}$, is to increase $\mathcal{J}(\mathfrak{S}(z))$ (i.e., improve the quality of fake bills) by observing $\mathfrak{D}$'s differentiation process. In the process of "busting fake bills" and "making better fake bills", the model distribution $p_{\mathfrak{D}}$ approaches p_{data} and eventually reaches an equilibrium, where $\mathfrak{D}$ no longer excels over chance (i.e., $\mathfrak{D}(x) = \mathfrak{D}(\mathfrak{S}(z)) = 0.5$). Consequently, G reaches its optimal point [40] during counterfeiting.

4.2. Objective Function

GANs are unsupervised models based on game theory, proposed by Goodfellow et al. in 2014. In this model, the generator $\mathfrak{S}$ maps random noise onto the target domain with a particular distribution (such as Gaussian or uniform), and the probability distribution of the real data is then studied in order to generate samples $\mathfrak{S}(z)$ that closely match the real data distribution. To avoid the mode collapse problem when random noise is used, $\mathfrak{D}$ identifies whether the sample derives from the real dataset x or from the generated dataset $\mathfrak{S}(z)$. $\mathfrak{D}(\cdot)$ is a probability value that implies whether the dataset is real. Training a GAN primarily involves training the discriminator $\mathfrak{D}$ and generator $\mathfrak{S}$, where $\mathfrak{S}(z)$ is trained to reduce the likelihood of $\mathfrak{D}$ making mistakes, whereas $\mathfrak{D}$ is used to increase the precision

level for distinguishing real from generated samples [17]. Hence, training $\mathfrak{S}(z)$ and $\mathfrak{D}$ is a binary minimax game problem that is characterized by the following objective function:

$$J(\mathfrak{D}, \mathfrak{S}) = \mathbb{E}_{x \sim p_{data}}[\log \mathfrak{D}(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - \mathfrak{D}(\mathfrak{S}(z)))] \quad (1)$$

where $x \sim p_{data}$ implies that datum x follows some p data distribution, and $z \sim p_z$ implies that the random noise z follows the distribution p_z.

When the generator trains a GAN, it attempts to generate progressively realistic images so that a discriminator will not be able to distinguish between real and false images, whereas $\mathfrak{D}$ attempts to distinguish between real images and those generated by $\mathfrak{S}$. This is similar to a two-player game in which the generator and the discriminator are steadily at odds. Eventually, a dynamic equilibrium is reached between the two networks; the image generated by $\mathfrak{S}$ represents the real image, whereas the discriminator cannot distinguish between them. As a result, network training is ultimately intended to maximize the probability values of the $\mathfrak{S}$ and $\mathfrak{D}$ networks, which means that $\mathfrak{S}$ and $\mathfrak{D}$ should achieve "Nash equilibrium".

In order to stabilize the training of GANs, different objective functions have been proposed, including f-divergence [41], least squares [42], Wasserstein distance [43], and hinge loss [44]. The Wasserstein metric is feasibly the most favored for the measurement of distances between real and generated samples, developing a new GAN framework called the Wasserstein GAN (WGAN) [43]. A significant advantage of the WGAN over the original GAN is its improved theoretical properties and learning stability, which facilitate hyperparameter selection and debugging. Gulrajani et al. [45] improved the original WGAN by penalizing the norm of the discriminator gradients with respect to the data input rather than clipping the weight parameters of the discriminator.

It is often possible to parameterize networks $\mathfrak{D}$ and G according to differentiable functions that may include fully connected, convolutional, or recurrent networks. In the original GAN [17], fully connected networks served as the building blocks, which were suited only to low-resolution image datasets such as MNIST and CIFAR-10. Increasingly powerful convolutional neural networks (CNNs) have been used to synthesize high-resolution complex images, providing better image generation performance. Several GAN variants have been proposed in the literature based on the vanilla GAN model [17] for various applications [46–48].

By conditioning the GAN on auxiliary information, it can be modified to generate images with desired attributes—as opposed to the original GAN, which has no control over the data being generated. A conditional GAN (cGAN) is a type of artificial neural network conditioned on a class label. In this study, we aim to predict future floral scent compounds of Taif rose based on GC-MS fingerprints acquired during the distillation process using a conditional GAN (cGAN) [33]. The chemical components of rose oil are referred to as the visible phenotypes of plants at different locations. Using GC–MS fingerprint analysis of 10 Taif rose samples, we generated images of the future phytoconstituents in rose oil. Mirza and Osindero [22] were among the first researchers to develop conditional GANs (cGANs). Using z and y as inputs, the $\mathfrak{S}$ represents random noise in the joint representation, and $\mathfrak{D}$ represents real samples and their labels. A cGAN has the following objective function:

$$J(\mathfrak{D}, \mathfrak{S}, y) = \mathbb{E}_{x \sim p_{data}}[\log \mathfrak{D}(x|y)] + \mathbb{E}_{z \sim p_z}[\log(1 - \mathfrak{D}(\mathfrak{S}(z|y)))] \quad (2)$$

4.3. Network Design (Architecture)

Using user-generated multilabel predictions, Mirza and Osindero [22] developed a class-conditional GAN that generated MNIST digits with class labels encoded as one-hot vectors. The auxiliary classifier GAN (or AC-GAN) performs the cGAN function by predicting class labels rather than taking them as inputs [21], leading to dispersion and diversity in the ImageNet samples. Furthermore, the conditional variable y may include images [20,33], bounding boxes, key points, or images with text descriptions. According to Reed et al. [40], 64×64 plausible images can be generated if textual descriptions are

available. StackGANs were proposed by Zhang et al. [35], which produced images with 4X better resolution for text-to-image synthesis. Isola et al. [33] proposed Pix2Pix (https://github.com/phillipi/pix2pix, accessed on 15 February 2023) for image-to-image translation (i.e., translation of representations from the source to output images). To learn the mappings between the input and output images, Pix2Pix requires a set of aligned image pairs as the training data. The newly introduced image-conditional GAN has been widely applied in image synthesis, face editing, image inpainting, and super-resolution [47,49,50]. Figure 3 summarizes the key parameters of the architecture used in this study.

4.3.1. The Generator

The generator network begins with two fully connected layers, whose number of neurons is set to 1000 and 10,000, respectively. The main task of these two layers is to convert the dimensional reduction sample input to the network into two higher dimensions so that subsequent layers can use them to reproduce a new set of samples. The outputs of the two fully connected layers are then passed to five convolutional nonlinearity tuples as the number of kernels and the width of the convolutional layers vary among the layers. This is made with the aim of enabling the kernel of these layers to discover different parts of the patterns provided in the given samples—which, in turn, increases the possibility of generating new samples with the same patterns. The final layer of the generator network is a single convolutional layer, and the number and size of its kernels are set to match the number and size of the frontmost convolutional layer to ensure that the generated input is identical to the input.

4.3.2. The Discriminator

The discriminator network consists of a tuple of five, each comprising "convolutional nonlinearity and max-pooling followed by three fully connected layers". Recall that the main goal of the discriminator network is to determine whether each input is genuine or generated by the generative network. To achieve this goal, five tuples are used to extract the salient features of the inputs, whereas the fully connected layers are used to embed these features in the learning space before passing them on to the last layer, which consists of two neurons representing the two possible cases: genuine, or generated by the generative network.

The number of stacks refers to the number of generator and discriminator networks used in this study, whereas the distribution of p_y refers to the probability distribution with which a new sample is drawn from the dataset and fed into the model. The greatest advantage of using a uniform distribution is that all the samples have the same chance of being used. Furthermore, a Gaussian distribution with a mean of zero and a small standard deviation is used to create random noise for the input. The choice of this distribution is justified by the fact that almost all noise occurring in the real field follows this distribution.

The proposed model was implemented on a single machine running Ubuntu 22.04.1 Long-Term Support with an Intel® Core™ i9-10900X X-series processor (19.25M cache, 3.70 GHz), 64 GB, and a GeForce RTX 2060 graphics card. Python 3.10.7, with the TensorFlow and Keras libraries, was used as the programming environment, over which the model was coded.

4.4. Validation

A cluster model was developed using the dataset provided to quantify the diversity that must be implemented in the proposed model. Cross-entropy and minimax loss functions were used to train the discriminator and generator networks, respectively. An 80:20 ratio was used to divide the datasets into two subsets: training and testing. Adaptive moment estimation with $\beta 1 = 0.995$ and $\beta 2 = 0.99$ was used as an optimization algorithm, while backpropagation was used as a learning scheme with a learning rate of 0.001 and 600 epochs.

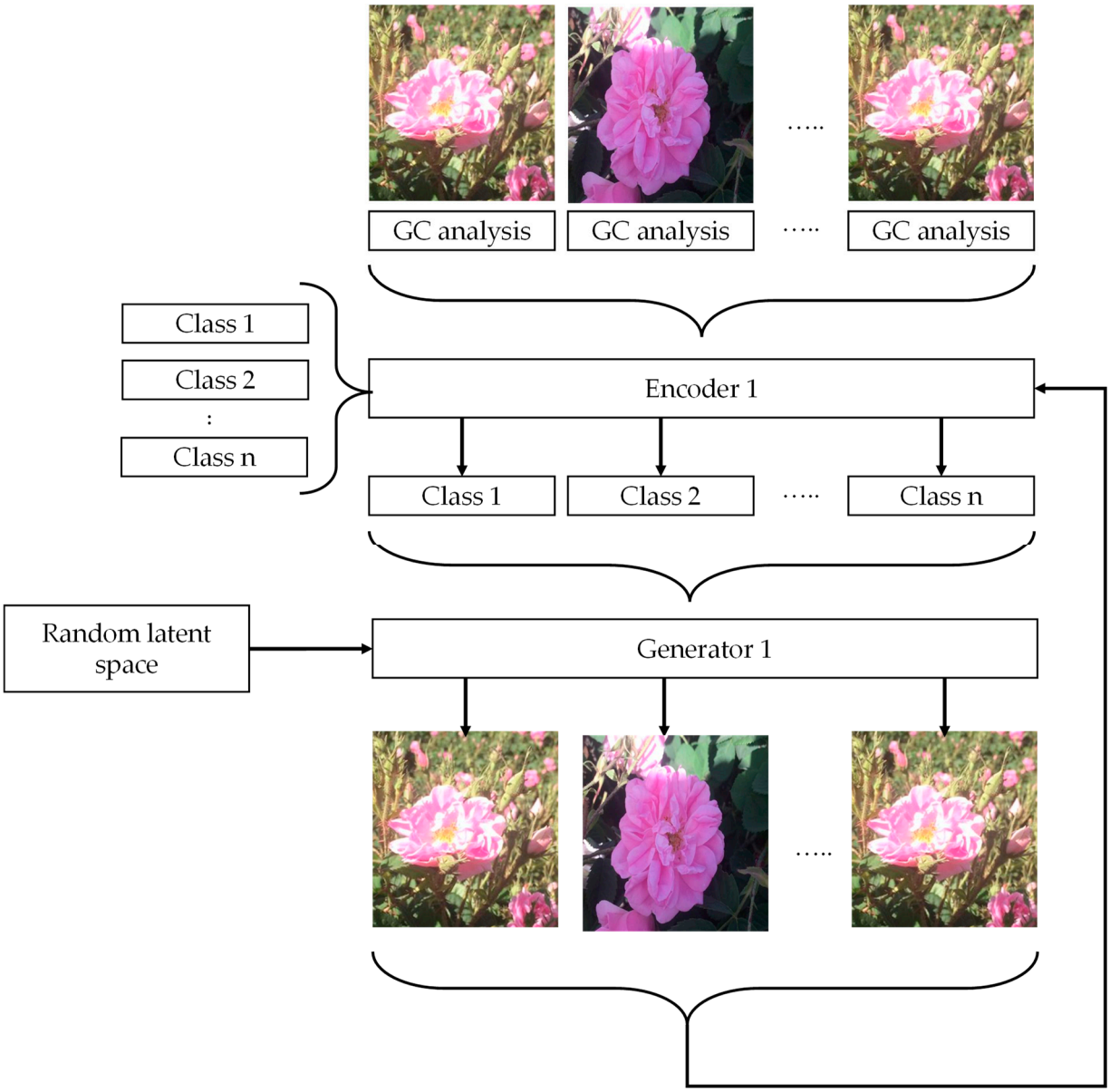


Figure 3. A schematic diagram of the proposed cSGAN model used in this study.

5. Results

The first evaluation of the proposed model was conducted by measuring the suitability of the proposed approach for generating new samples from a given dataset. In this evaluation, the retention time (RT) was measured against the molecular weight (MW) for different molecular formulae (MFs) in terms of area and then plotted as shown in Figure 4. It can be seen that the different accessions of Taif rose have variable metabolomic profiles. For example, those collected from location 3 show that most of the components are concentrated below 400 MW and have a lower percentage area, whereas this is twice the case for site 2. The most important point to consider when developing the model is to provide sufficient layers to account for the different distributions in the dataset.

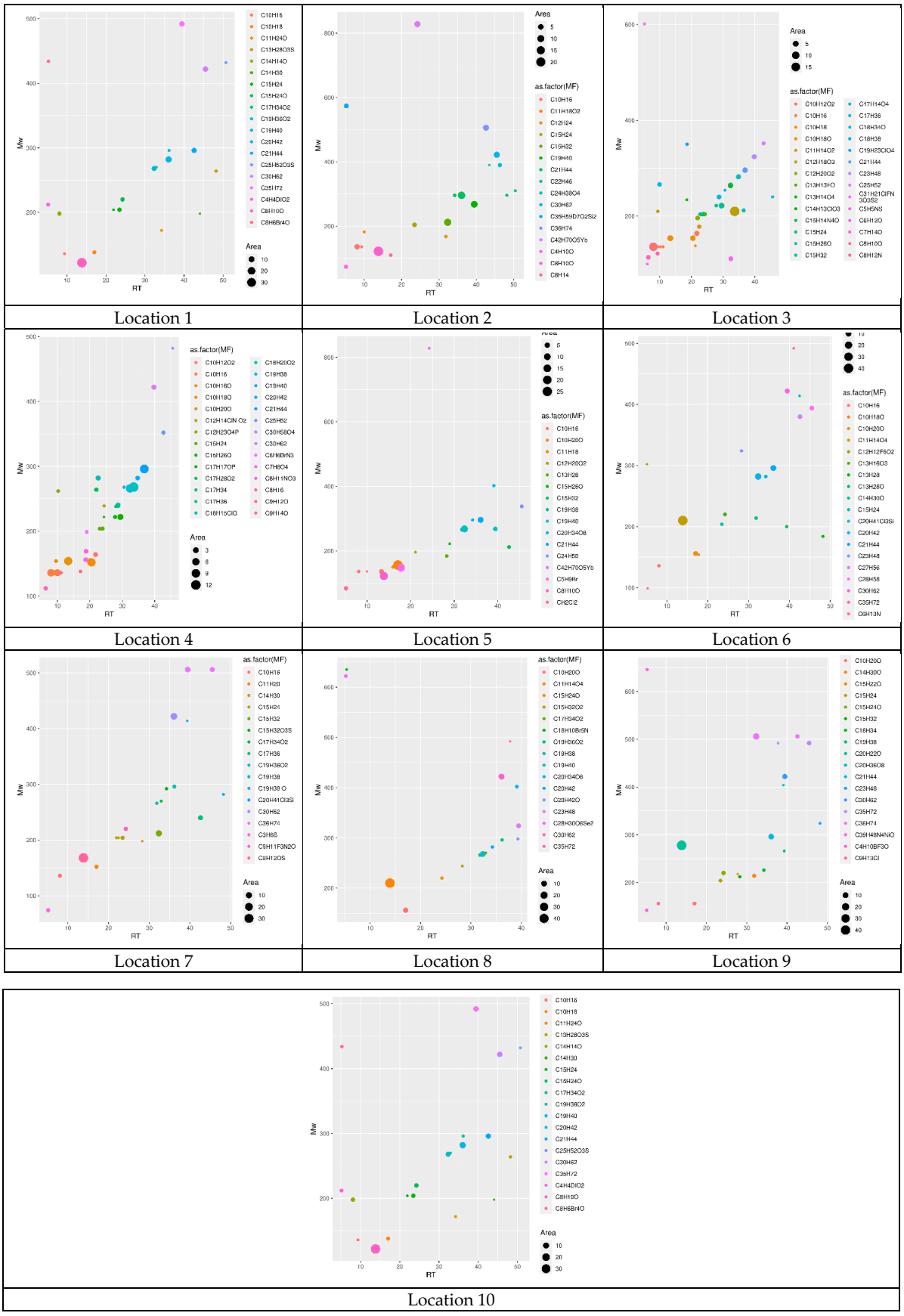


Figure 4. Bubble figures of RT vs. Mw for all of the locations. It is clear that Taif rose accessions are different in their oil composition characteristics.

To further quantify the diversity that must be implemented in the proposed model, a cluster model was created for the provided dataset and used to assess the extent to which the different aspects of the datasets matched. The results of this assessment, as shown in Figure 5A,B, indicate that it is difficult to group all sites into three or four clusters. The optimal number of clusters, as shown in the cluster gap analysis in Figure 5B, indicates that an optimal value is achieved when each location is placed in its own cluster.

The accuracy of the proposed model was assessed by measuring losses as a function of the number of epochs. From Figure 6, it can be observed that as the number of epochs increased, the losses for both the generator and discriminator networks decreased during the training and testing phases. This result demonstrates the ability of the proposed model to perfectly handle a dataset.

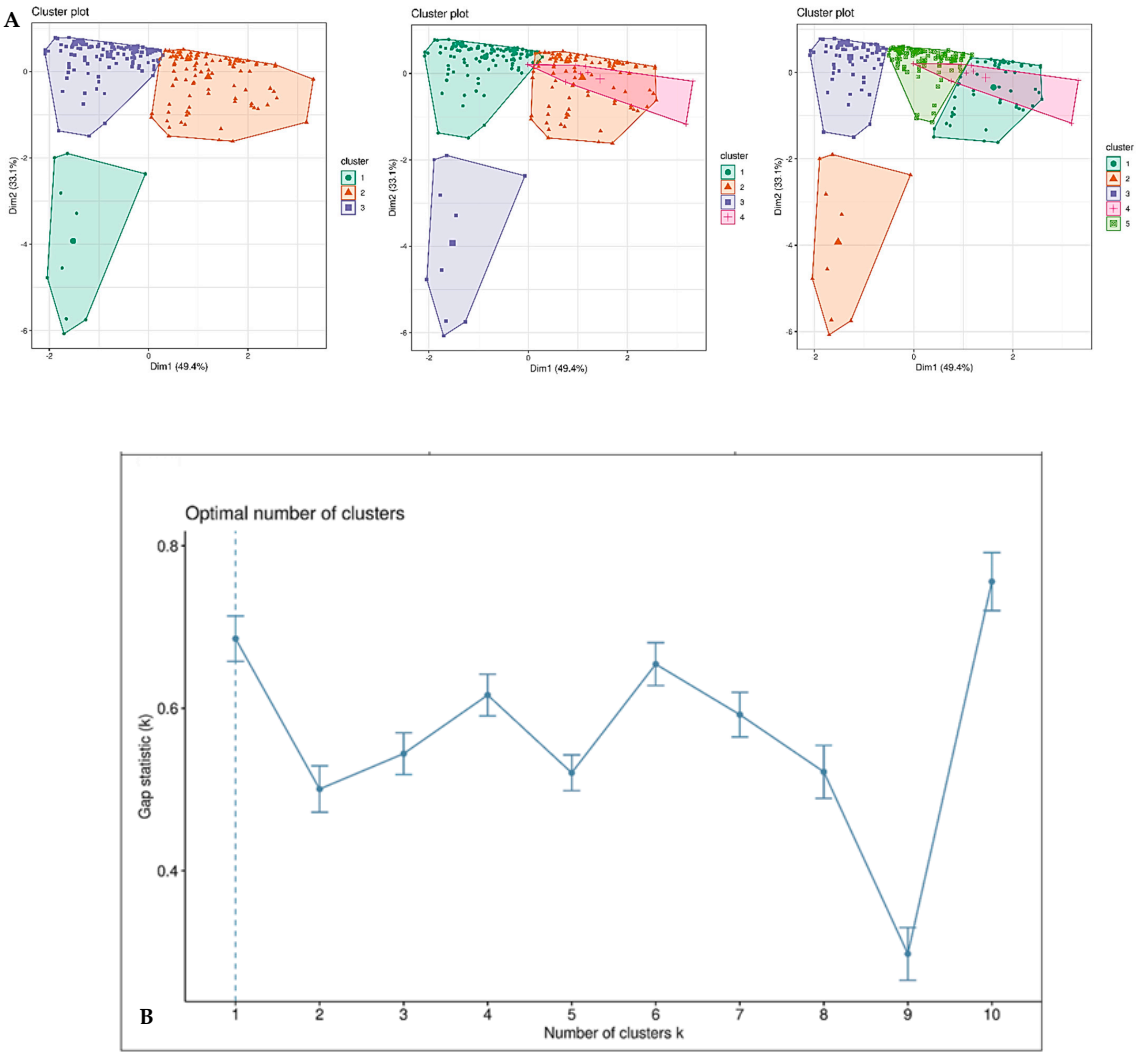


Figure 5. (**A**) Cluster analysis of the dataset provided and an assessment of how closely the different aspects of the datasets are matched. (**B**) Cluster gap analysis clarifies the optimal number of clusters and indicates that the optimal value is achieved when each location is placed in a separate cluster.

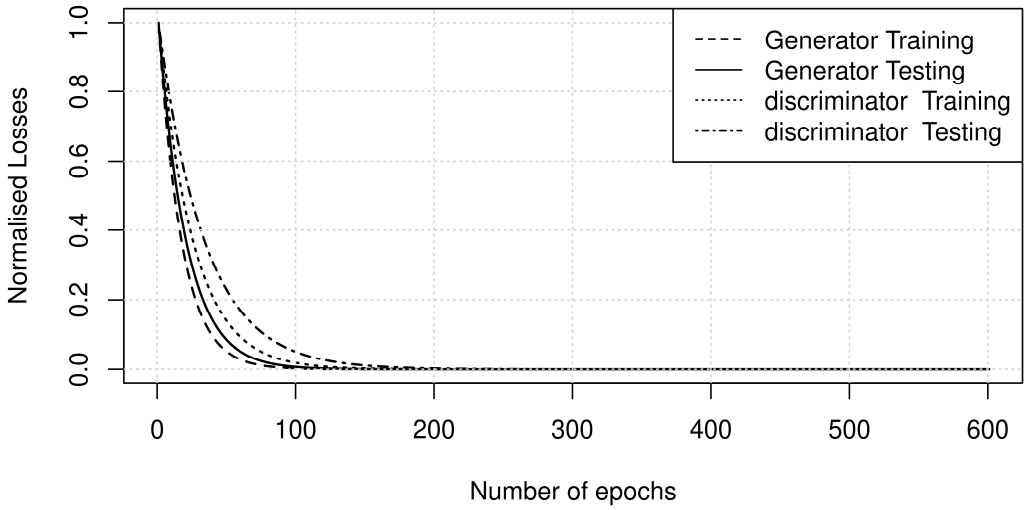


Figure 6. Graph of loss function changes versus number of epochs for the generator and discriminator during training.

6. Discussion

Synthetic data generation poses the challenge of capturing the complex patterns of real data. To determine the accuracy and quality of the synthetic data, a rigorous evaluation method must be used, followed by a direct comparison of the synthetic data produced by various generative models. It is noteworthy, however, that different evaluation strategies and metrics result in different results for generative models with high heterogeneity and complexity [51], particularly for high-dimensional data. In addition, the results of evaluation may differ based on the application domain and may not be simply transferable to other applications. In order to address this current target problem, a specific evaluation framework must be developed.

Our study is primarily concerned with agricultural applications, for which no real data exist. Therefore, we introduced an extended dataset including both generated and real data to train the classifier. As a result, we can see how synthesized data can improve the accuracy of the classifier, while at the same time retaining the features of real data. Comparatively, other evaluation methods require the classifier to be trained exclusively on real data and to be tested on synthetic data [52,53]. On the other hand, other researchers used synthetic data to train their classifiers and then tested them on real data [54].

It should be noted that, in the present study, the same real data were employed to train the generative models, and partially in an expanded dataset to train the classifier. However, this allows us to address one of the biggest challenges in agriculture—namely, data scarcity, where data are limited. Synthetic data are important because they can overcome the data limitations arising from a variety of factors that lead to complications in the field. However, it should be noted that these data are typically of lower quality than real data, as confirmed by supervised classification accuracy [55]. Based on the available (limited) real data, the amount of synthetic data required must be determined in order to ensure high accuracy. This study proposes a generative model called the conditional stacked generative adversarial network (cSGAN). It comprises a top-down stack of GANs that generate a set of "plausible" lower-level representations based on higher-level representations. Similarly, we introduced a set of representation discriminators, which are trained to distinguish "fake" representations from "real" representations, similar to the image discriminator in the original GAN model. The loss created by the representation discriminator drives the intermediate representations of the SGAN to lie on the manifold of the representation space of the bottom-up DNN. In this regard, the proposed GAN variant (i.e., StackGAN) could improve classification accuracy compared to a vanilla GAN trained only on real data [53]. Generative models encounter problems when a limited real data are considered, resulting in the generation of inferior data. This is particularly problematic in agriculture,

where only limited amounts of data are available. Consequently, generating and testing data quality is difficult, as opposed to studies that use large datasets [53]. It is often used for clustering, where data points are grouped based on their similarity, which allows the detection of unknown correlations and patterns in the datasets. Clustering is a common application of unsupervised learning (UL), as it involves grouping data points according to their similarity, allowing the discovery of unknown correlations and patterns in the data. Another important task of UL is dimensionality reduction, which aims to find efficient low-dimensional representations of data while minimizing the resulting loss of information. This is closely related to finding a cluster of data with similar features. A popular and relatively simple clustering algorithm is the k-means method [56], for which a variety of variations and extensions have been proposed. For example, the k-means algorithm minimizes the average variance in the clusters.

It is worth mentioning that each model is run with a different number of iterations, and the number of epochs is an influential parameter. When the model is run with fewer iterations, the accuracy is low and the error is larger. As the number of iterations increases, the model gradually converges such that there is no significant difference between 300 and 500 epochs [57]. In this context, it was determined that a maximum of 600 epochs is sufficient. The number of neurons in the hidden layers is another parameter that affects the accuracy of the models. A low value prevents the model from simulating correctly, whereas a high value leads to overfitting. The loss was measured as a function of the number of epochs in the proposed model to evaluate its accuracy. In both the training and testing phases, the losses for both the generator and discriminator networks increased as the number of epochs increased.

The graphs of the loss function show that, for both the generator and the discriminator, the error in the test phase decreased as the modelling error in the training phase decreased. Furthermore, the distance between the two lines in the graph decreased. Thus, the dropout function was effective in preventing overfitting of the network. A well-chosen learning rate was also ensured by examining the changes in the loss function value over epochs. From this result, it can be concluded that the proposed model can perfectly process the dataset.

7. Conclusions

The primary aim of this study was to develop a general prediction model for Taif rose populations using recently improved generative adversarial networks. According to the findings of this study, the cSGAN model could produce artificial samples from 10 Taif rose samples collected from various locations. GC-MS metabolite profiling of sparse Taif rose populations generated by cGANs leads to improved downstream analyses such as detecting marker metabolites, testing novel analysis algorithms, and assessing the robustness and reliability of classifiers. This would lead to fewer plant experiments and lower costs. The results of the present study indicate that the cGAN model has great potential and value. It is only a short time since cGAN has existed; therefore, relevant theories and applications are still in their infancy. Further research on cGAN is needed to consolidate its development, and this work could hopefully provide information that researchers need in order to apply cGAN to agricultural issues.

Author Contributions: Conceptualization, H.M.A. and M.B.; methodology, M.B. and M.A.; software, H.G.Z.; validation, M.B., M.A.A. and J.F.A.-A.; formal analysis, M.B.; investigation, M.A.A., N.A.A. and M.A.; resources, M.M.M. and M.A.; data curation, M.B.; writing—original draft preparation, M.B. and M.A.; writing—review and editing, H.M.A.; visualization, M.A.A.; supervision, H.M.A.; project administration, H.M.A.; funding acquisition, H.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the research project No. [1-441-126] from the Ministry of Education in Saudi Arabia.

Acknowledgments: The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, for funding this research work through project number [1-441-126].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. USDA, NRCS. "Rosa × damascena". The PLANTS Database. National Plant Data Team, Greensboro, NC 27401-4901 USA. Available online: http://plants.usda.gov (accessed on 22 January 2023).
2. Rusanov, K.E.; Kovacheva, N.M.; Atanassov, I.I. Comparative GC/MS Analysis of Rose Flower and Distilled Oil Volatiles of The Oil Bearing Rose. *Rosa Damascena. Biotechnol. Biotechnol. Equip.* **2011**, *25*, 2210–2216. [CrossRef]
3. Rusanov, K.; Kovacheva, N.; Stefanova, K.; Atanassov, A.; Atanassov, I. Rosa damascena—Genetic resources and capacity building for molecular breeding. *Biotechnol. Biotechnol. Equip.* **2009**, *23*, 1436–1439. [CrossRef]
4. Rusanov, K.; Kovacheva, N.; Vosman, B.; Zhang, L.; Rajapakse, S.; Atanassov, A. Microsatellite analysis of Rosa damascena Mill. accessions reveals genetic similarity between genotypes used for rose oil production and old Damask rose varieties. *Theor. Appl. Genet.* **2005**, *111*, 804–809. [CrossRef] [PubMed]
5. Rusanov, K.; Kovacheva, N.; Rusanova, M.; Atanassov, I. Flower phenotype variation, essential oil variation and genetic diversity among Rosa alba L. accessions used for rose oil production in Bulgaria. *Sci. Hortic.* **2013**, *161*, 76–80. [CrossRef]
6. ISO 9842:2003; International Standards for Business, Government and Society. Available online: www.iso.org (accessed on 15 November 2022).
7. Krishnan, B. Classification Of Chemicals Present In Essential Oils Using Deep Learning Algorithm. *Nveo-Nat. Volatiles Essent. Oils J. NVEO* **2021**, *8*, 3607–3617.
8. Kim, H.U.; Sohn, S.B.; Lee, S.Y. Metabolic network modeling and simulation for drug targeting and discovery. *Biotechnol. J.* **2012**, *7*, 330–342. [CrossRef] [PubMed]
9. Zhang, X.; Acencio, M.L.; Lemke, N. Predicting essential genes and proteins based on machine learning and network topological features: A comprehensive review. *Front. Physiol.* **2016**, *7*, 1–11.
10. Saa, P.A.; Nielsen, L.K. Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach. *Sci. Rep.* **2016**, *6*, 29635. [CrossRef]
11. Zampieri, G.; Vijayakumar, S.; Yaneske, E.; Angione, C. Machine and deep learning meet genome-scale metabolic modeling. *PLOS Comput. Biol.* **2019**, *15*, e1007084. [CrossRef] [PubMed]
12. Pisner, D.A.; Schnyer, D.M. Support vector machine. In *Machine Learning*; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121.
13. Vijayakumar, S.; Conway, M.; Lió, P.; Angione, C. Seeing the wood for the trees: A forest of methods for optimization and omic-network integration in metabolic modelling. *Brief. Bioinform.* **2017**, *19*, 1218–1235. [CrossRef] [PubMed]
14. Alakwaa, F.M.; Chaudhary, K.; Garmire, L.X. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J. Proteom. Res.* **2018**, *17*, 337–347. [CrossRef] [PubMed]
15. Pomyen, Y.; Wanichthanarak, K.; Poungsombat, P.; Fahrmann, J.; Grapov, D.; Khoomrung, S. Deep metabolome: Applications of deep learning in metabolomics. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2818–2825. [CrossRef] [PubMed]
16. Lawson, C.E.; Martí, J.M.; Radivojevic, T.; Jonnalagadda, S.V.R.; Gentz, R.; Hillson, N.J.; Peisert, S.; Kim, J.; Simmons, B.A.; Petzold, C.J.; et al. Machine learning for metabolic engineering: A review. *Metab. Eng.* **2020**, *63*, 34–60. [CrossRef]
17. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Bing, X.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the International Conference on Neural Information Processing Systems, Kuching, Malaysia, 3–6 November 2014.
18. Lin, Y.; Dai, X.; Li, L.; Wang, X.; Wang, F. The new Frontier of AI research: Generative adversarial networks. *Acta Autom. Sin.* **2018**, *44*, 775–792. [CrossRef]
19. Hiasa, Y.; Otake, Y.; Takao, M.; Matsuoka, T.; Sato, Y. Cross-modality image synthesis from unpaired data using CycleGAN: Effects of gradient consistency loss and training data size. In *Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018*; Goksel, O., Oguz, I., Gooya, A., Burgos, N., Eds.; Springer: Granada, Spain, 2018.
20. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Odena, A.; Olah, C.; Shlens, J. Conditional Image Synthesis with Auxiliary Classifier GANs. In Proceedings of the ICML 2017, Sydney, NSW, Australia, 6–11 August 2017; pp. 2642–2651.
22. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
23. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In *International Conference on Information Processing in Medical Imaging*; Springer: Cham, Switzerland, 2017; pp. 146–157.
24. Zenati, H.; Foo, C.S.; Lecouat, B.; Manek, G.; Chandrasekhar, V.R. Efficient gan-based anomaly detection. In The Workshop on International Conference on Learning Representations. *arXiv* **2018**, arXiv:1802.06222.

25. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial feature learning. In International Conference on Learning Representations. *arXiv* **2017**, arXiv:1605.09782.
26. Akçay, S.; Atapour-Abarghouei, A.; Breckon, T.P. GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. In *Computer Vision—ACCV 2018*; Springer: Cham, The Netherlands, 2018; pp. 622–637.
27. Deecke, L.; Vandermeulen, R.; Ruff, L.; Mandt, S.; Kloft, M. Image anomaly detection with generative adversarial networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: Cham, Switzerland, 2019; pp. 3–17.
28. Soleimani, H.; Miller, D.J. ATD: Anomalous Topic Discovery in High Dimensional Discrete Data. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 2267–2280. [CrossRef]
29. Aggarwal, C.C. An introduction to outlier analysis. In *Outlier Analysis 2017*; Springer: Cham, Switzerland, 2017; pp. 1–34.
30. Yang, X.; Latecki, L.J.; Pokrajac, D. Outlier Detection with Globally Optimal Exemplar-Based GMM. In Proceedings of the 2009 SIAM International Conference on Data Mining, Sparks, NA, USA, 30 April–2 May 2009.
31. Zong, B.; Song, Q.; Min, M.R.; Cheng, W.; Lumezanu, C.; Cho, D.; Chen, H. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In Proceedings of the 6th International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
32. Drees, L.; Junker-Frohn, L.V.; Kierdorf, J.; Roscher, R. Temporal prediction and evaluation of Brassica growth in the field using conditional generative adversarial networks. *Comput. Electron. Agric.* **2021**, *190*, 106415. [CrossRef]
33. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976.
34. Hong, Y.; Zhou, W.; Zhang, J.; Zhou, G.; Zhu, Q. Self-regulation: Employing a generative adversarial network to improve event detection. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Cedarville, OH, USA; Volume 1: Long Papers, pp. 515–526.
35. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D.N. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5907–5915.
36. Antipov, S.S.; Tutukina, M.N.; Preobrazhenskaya, E.V.; Kondrashov, F.A.; Patrushev, M.V.; Toshchakov, S.V.; Dominova, I.; Shvyreva, U.S.; Vrublevskaya, V.V.; Morenkov, O.S.; et al. The nucleoid protein Dps binds genomic DNA of Escherichia coli in a non-random manner. *PLoS ONE* **2017**, *12*, e0182800. [CrossRef] [PubMed]
37. Xiao, Z.; Luo, J.; Niu, Y.; Wu, M. Characterization of key aroma compounds from different rose essential oils using gas chromatography-mass spectrometry, gas chromatography–olfactometry and partial least squares regression. *Nat. Prod. Res.* **2017**, *32*, 1567–1572. [CrossRef] [PubMed]
38. National Institute of Standards and Technology. *Security Requirements for Cryptographic Modules*; Technical Report Federal Information Processing Standards Publications (FIPS PUBS) 140-2; U.S. Department of Commerce: Washington, DC, USA, 2022.
39. Goodfellow, I. Nips 2016 tutorial: Generative adversarial networks. *arXiv* **2016**, arXiv:1701.00160.
40. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. Generative adversarial text to image synthesis. *arXiv* **2016**, arXiv:1605.05396.
41. Nowozin, S.; Cseke, B.; Tomioka, R. f-gan: Training generative neural samplers using variational divergence minimization. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 271–279.
42. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.
43. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. In International conference on machine learning 2017 Jul 17 (pp. 214–223). PMLR. *arXiv* **2017**, arXiv:1701.07875.
44. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
45. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5769–5779.
46. Wang, R.; Bashyam, V.; Yang, Z.; Yu, F.; Tassopoulou, V.; Sreepada, L.P.; Chintapalli, S.S.; Sahoo, D.; Skampardoni, I.; Nikita, K.; et al. Applications of Generative Adversarial Networks in Neuroimaging and Clinical Neuroscience. *arXiv* **2022**, arXiv:2206.07081. [CrossRef] [PubMed]
47. Huang, X.; Liu, M.Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV) 2018, Munich, Germany, 8–14 September 2018; pp. 172–189.
48. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
49. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
50. Gui, G.; Liu, M.; Tang, F.; Kato, N.; Adachi, F. 6G: Opening New Horizons for Integration of Comfort, Security, and Intelligence. *IEEE Wirel. Commun.* **2020**, *27*, 126–132. [CrossRef]
51. Theis, L.; Oord, A.v.d.; Bethge, M. A Note on the Evaluation of Generative Models. *arXiv* **2015**, arXiv:1511.01844.
52. Goncalves, A.; Ray, P.; Soper, B.; Stevens, J.; Coyle, L.; Sales, A.P. Generation and evaluation of synthetic patient data. *BMC Med. Res. Methodol.* **2020**, *20*, 108. [CrossRef] [PubMed]

53. Zhao, Z.; Kunar, A.; Birke, R.; Chen, L.Y. CTAB-GAN: Effective Table Data Synthesizing. In Proceedings of the 13th Asian Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 97–112.
54. Xu, L.; Skoularidou, M.; Cuesta-Infante, A.; Veeramachaneni, K. Modeling Tabular data using Conditional GAN. *arXiv* **2019**, arXiv:1907.00503.
55. Rankin, D.; Black, M.; Bond, R.; Wallace, J.; Mulvenna, M.; Epelde, G. Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing. *JMIR Med. Inform.* **2020**, *8*, e18910. [CrossRef]
56. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
57. Apaydin, H.; Feizi, H.; Sattari, M.T.; Colak, M.S.; Shamshirband, S.; Chau, K.-W. Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. *Water* **2020**, *12*, 1500. [CrossRef]