**ARTICLE**     OPEN

Check for updates

# GestaltGAN: synthetic photorealistic portraits of individuals with rare genetic disorders

Aron Kirchhoff [1], Alexander Hustinx [1], Behnam Javanmardi [1], Tzung-Chien Hsieh [1], Fabian Brand[1], Fabio Hellmann [2], Silvan Mertes[2], Elisabeth André[2], Shahida Moosa [3], Thomas Schultz [4], Benjamin D. Solomon [5] and Peter Krawitz[1] ✉

The facial gestalt (overall facial morphology) is a characteristic clinical feature in many genetic disorders that is often essential for suspecting and establishing a specific diagnosis. Therefore, publishing images of individuals affected by pathogenic variants in disease-associated genes has been an important part of scientific communication. Furthermore, medical imaging data is also crucial for teaching and training deep-learning models such as GestaltMatcher. However, medical data is often sparsely available, and sharing patient images involves risks related to privacy and re-identification. Therefore, we explored whether generative neural networks can be used to synthesize accurate portraits for rare disorders. We modified a StyleGAN architecture and trained it to produce artificial condition-specific portraits for multiple disorders. In addition, we present a technique that generates a sharp and detailed average patient portrait for a given disorder. We trained our GestaltGAN on the 20 most frequent disorders from the GestaltMatcher database. We used REAL-ESRGAN to increase the resolution of portraits from the training data with low-quality and colorized black-and-white images. To augment the model's understanding of human facial features, an unaffected class was introduced to the training data. We tested the validity of our generated portraits with 63 human experts. Our findings demonstrate the model's proficiency in generating photorealistic portraits that capture the characteristic features of a disorder while preserving patient privacy. Overall, the output from our approach holds promise for various applications, including visualizations for publications and educational materials and augmenting training data for deep learning.

## INTRODUCTION
Many genetic conditions involve features evident on physical examination, including those that affect the face. At the time of writing (May 28, 2024), searching with the HPO term "facial dysmorphism" yields 2997 entries in the Online Mendelian Inheritance of Men (OMIM) compendium (https://www.omim.org/), indicating the importance of the facial gestalt for characterizing disease entities. The importance of phenotype matching extends to genetic diagnostic procedures, where physical examination features can serve as supporting evidence when assessing sequence variants for pathogenicity [1].

Recent advancements in computer vision have achieved expert-level accuracy in discerning distinct facial patterns. Next-generation phenotyping (NGP) tools such as GestaltMatcher have become instrumental in analyzing clinical patterns in human faces and their usage for interpreting sequencing data [2–4]. The underlying technology, deep-learning, can be used for pattern recognition and delineating informative features (explainable AI, XAI) or synthesizing images with similar characteristics via generative methods such as Generative Neural Networks (GNN) [5]. GNNs may be particularly useful in medical settings since data are often sparsely available and may involve sensitive, private information. The generated images can be used for teaching or data augmentation when training machine learning models, including to address privacy concerns [6, 7]. In medical genetics, Duong et al. showed that StyleGAN could generate artificial longitudinal patient data and improve NGP classification accuracy by better controlling age as a confounder [8].

StyleGAN is now a well-established architecture for image generation that allows the synthesis of photorealistic images across diverse contexts [9]. StyleGAN is based on the concept of Generative Adversarial Networks (GAN) originally proposed by Goodfellow et al., which consists of two parts: the generator and the discriminator [10]. The generator crafts images—such as human portraits—while the discriminator evaluates their quality and originality, providing feedback to reduce artifacts and enhance realism. The generator's goal is to fool the discriminator with its results such that the discriminator cannot tell whether they are real or synthetic. Through this adversarial process, the generator learns characteristic object properties required to produce realistic synthetic images (in this case, human faces). A more comprehensive introduction to the technology can be found in the supplemental material (Related work).

With further refinement of GANs, it is possible to condition the output depending on an input label [11]. This label is an additional piece of information that enables one to conditionally generate a

certain type of image. In the case of human faces, the label might encode age, race or ethnicity, hair color, or even a certain genetic condition, as is the focus of our work.

Artificial content creation is particularly compelling in medicine, given the sparse availability and stringent privacy constraints on data. However, facial images are a particularly sensitive type of medical data, as the effort required for re-identification is relatively low and may require no additional technology. Nevertheless, for this study, the characteristics most suitable for de-identification can only be those that are not disease-related. Accordingly, there are limits to anonymization in so far as recognizing the disease is our aim (k-anonymity is bound by the prevalence of the disorder). However, sparse training data pose challenges, potentially leading to overfitting, a phenomenon where the network memorizes samples and recreates training images [7]. Balancing de-identification with feature retention poses a nuanced challenge; the model must learn and reproduce disorder-specific features without replicating exact facial combinations from training images.

We used "disorder" as an additional class label in our work. We trained a conditional StyleGAN with images of the GestaltMatcher database (GMDB) containing images of over 10,000 individuals with molecularly confirmed diagnoses [12]. We hypothesized that working with several syndromic disorders facilitates learning certain clinical features often shared by more than one disorder [12]. Therefore, we focused on the 20 most frequent syndromes represented in the GMDB, comprising a total of 3209 images. In order to enrich the characteristic features, we added a custom loss to the training and penalized the model if GestaltMatcher-Arc's [13] feedback would not match the syndrome requested of our GAN.

For the evaluation of the generated images, we tested whether humans are still able to differentiate between synthetic and original images, and whether the characteristic features of a genetic condition are preserved, while data of real patients is protected.

## METHODS AND MATERIALS
### Data preparation
The GestaltMatcher DataBase (GMDB) contains a collection of 581 distinct disorders known to involve facial dysmorphisms, with over 10,980 accompanying images of 8346 affected individuals. In addition to previously published images, all individuals newly represented in GMDB provided consent for use of their imaging data for machine learning. Ethical approval for the GMDB was granted by the IRB of the University Hospital Bonn. For our GestaltGAN model training, we focused on the 20 most common disorders from the GMDB (Supplemental Fig. 1). The reason for choosing 20 is a trade-off, balancing the benefit of using more disorders for training data against the challenge of distinguishing between them.

Images in the GMDB come in various formats, with differences in size, lighting, and facial alignment. To enhance the quality of low-resolution images, we used REAL-ESRGAN [14], a deep-learning model that predicts high-quality image details and computes high-resolution versions for the input images. Additionally, many older images in the GMDB come in black-and-white, which would lead to undesired outputs if directly used for training. We used DDColor [15] to add color to these monochromatic images to address this, ensuring a consistent dataset for our GAN training. All images were aligned and cropped using GestaltEngine-FaceCropper, which relies on RetinaFace [16] that accurately pinpoint five landmark points in each portrait: the eyes, nose, and mouth corners. Using those landmarks, horizontal alignment of the faces could be ensured.

Given the scarcity of images of individuals with the genetic conditions of interest, we expanded our dataset by including images of individuals without known genetic conditions. These unaffected faces share similar features, like hair or skin, with the images of individuals with genetic conditions, aiding the model in generating more realistic faces of individuals with genetic conditions. We opted for the FFHQ-Aging dataset [17], known for its size and high-quality images across different ages, races, and ethnicities. Since many patients in the GMDB are children (42.8%

under five years old), we balanced the age distribution by limiting the number of unaffected adults in FFHQ-Aging to 3000 individuals for all age groups over 20 years old. This adjustment resulted in a training set with 31,130 unaffected people. Subsequently, all images in FFHQ-Aging underwent alignment and cropping using GestaltEngine-FaceCropper.

## Training of GestaltGAN
We utilized the conditional StyleGAN3-R architecture proposed by Karras et al. [18], representing the fourth iteration of the StyleGAN framework. This version incorporates enhancements such as translation and rotation invariance towards training images, which helps with imperfect alignment that may have persisted in some training images. Notably, StyleGAN3-R integrates adaptive discriminator augmentation (ADA), a mechanism crucial for preventing overfitting, especially in datasets with limited samples like the GMDB.

Our approach is to use a conditional setup, where each syndrome is treated as a distinct class for training. This way, we can make the most of the shared facial features among different disorders using the size of our dataset. We also include unaffected images as a separate class, providing the model with continuous exposure to common features like hair or skin from those images. However, to handle the variable number of training images for each genetic condition and to avoid producing images that inappropriately incorporate features of unaffected faces, we added an over-sampling function to the StyleGAN3 implementation. This ensures that each genetic condition is represented equally during training. We gave the model twenty times more exposure to the unaffected class to allow the model to better understand and utilize its features.

An unaffected class was introduced to the training data to augment the model's understanding of human facial features. In addition to over-sampling, we modified the loss function of StyleGAN3. Since the GestaltMatcher model is specialized in identifying rare disorders from images, we want to leverage this skill by penalizing our model if its predictions deviate from what GestaltMatcher would expect for a given class. The adjusted loss function combines this GestaltMatcher loss with the regular discriminator loss $L_D$:

$$L_{total} = L_D(I) + \alpha \cdot \text{GestaltMatcherRank}(I, d)$$

Here, $I$ represents the image generated by the generator, and the GestaltMatcherRank function calculates the index of the correct disorder d has in the prediction of GestaltMatcher. The weight $\alpha$ adjusts the balance between the GestaltLoss and the discriminator loss.

The chosen image resolution for our model was $256 \times 256$ pixels, which is slightly below the median image resolution in the GMDB of $265 \times 328$ pixels. We were also able to train the model for a $512 \times 512$ resolution but did not continue this approach due to the three-fold higher required computation effort. Images generated by our GestaltGAN model can be seen in Fig. 1. A visualization of our training setup is shown in Fig. 2.

## Image averages and latent averages
To illustrate the characteristics of a disorder, the average face of several patients is often computed by registering and overlaying their portraits. While this method has been used in various studies, the resulting images are often blurry and indistinct [19]. Increasing the number of individuals often leads to a deterioration of the results (personal communications). In this subsection, we introduce a technique to generate sharp, high-quality portraits that accurately represent the features of specific disorders.

The image generation process of an image with StyleGAN begins with sampling a random latent vector $z \in Z$. This vector is combined with the class label in the mapping network, which maps to the second latent space $W$. An image is then deterministically generated based on this latent vector $w \in W$. Since the latent space is continuous, small variations in the latent vector result in slight variations in the synthesized image [20]. This means similar images, especially those of the same disorder, lie in the same region in the latent space. This property allows us to generate an average image for all disorders our model was trained on. To achieve this, we sample 10,000 latent vectors $w \in W$ for the selected disorder, such as Cornelia de Lange syndrome, and average these latent vectors to generate the average image. Instead of computing the average in the image space, we perform the averaging in the latent or feature space. Latent averages for different disorders are shown in Fig. 1 and Fig. 3. In Fig. 3, the latent averages are presented alongside the corresponding averages from image space, which are currently often used for teaching purposes, such as to help clinical trainees recognize different genetic conditions.
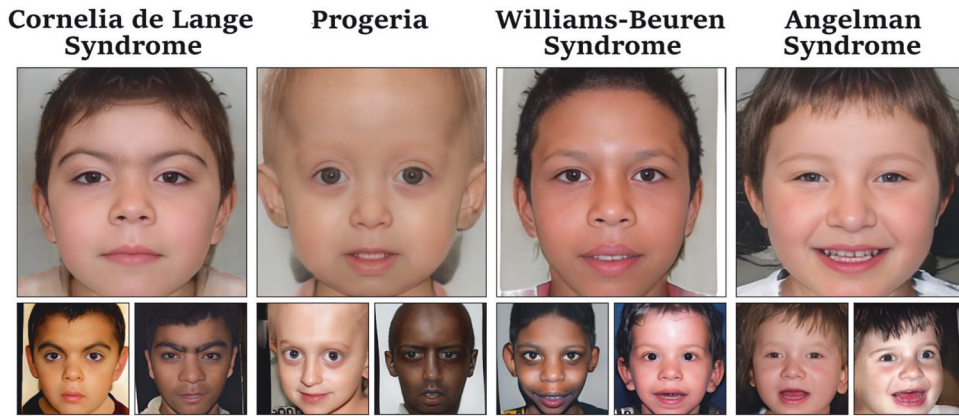
**Fig. 1 Images generated by GestaltGAN.** Images in the top row are the latent averages of the disorder, which were generated by averaging the features of the disorder. Images in the bottom row are selected images generated for the respective disorder.
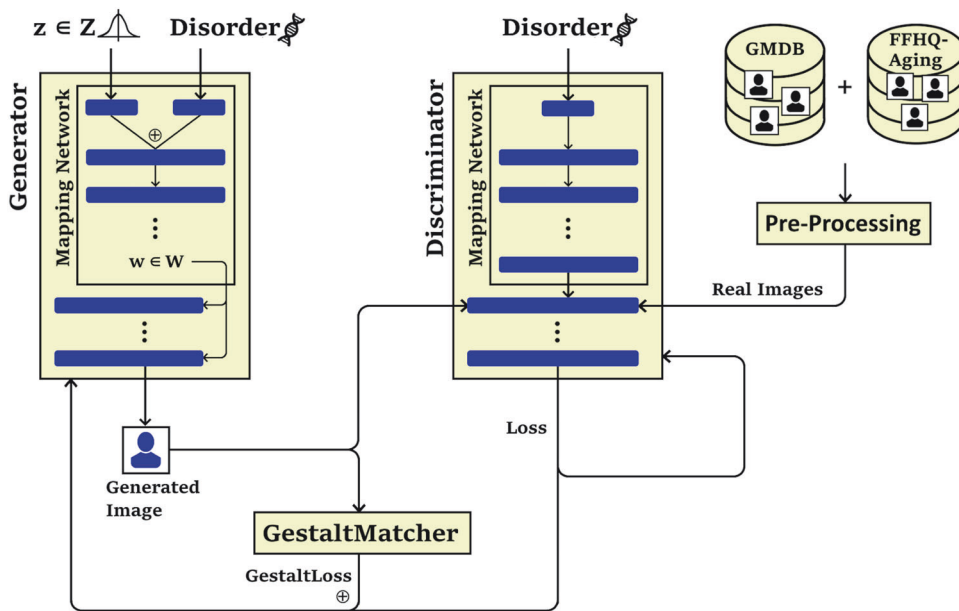


**Fig. 2 Visualization for the GestaltGAN architecture.** StyleGAN has been extended by a customized loss function, GestaltLoss, based on the GestaltMatcher ensemble. The conditional generator synthesizes images for 20 different disorders and receives feedback from the discriminator about the origin, which is either artificial or real. For the training of human faces and disorders, data of FFHQ Aging and the GMDB were used. Once the training is finished the generator can be used to synthesize arbitrary amounts of artificial images.

An important note is that this method only works for disorders on which the GAN has been trained. To extend the capability and generate latent averages for other disorders, we perform GAN-inversion on all patient portraits in the cohort to obtain the corresponding feature vectors in the latent space that represents the patients. We use the GestaltMatcher ensemble as a loss function for GAN inversion, as it has been trained to recognize dysmorphic features. After averaging the feature vectors, we generate an image from the resulting vector to obtain the latent average.

## RESULTS

With GestaltGAN, our goals were threefold: we aimed to create synthetic images that are photorealistic, de-identified, and accurately represent clinical features. Achieving these objectives partially involved navigating an optimization problem, as enhancing privacy protection might sometimes compromise feature preservation. Therefore, we assessed the quality of all three objectives through computational methods and by conducting experimental evaluations with human test participants, who compared the generated images to the original images. Images generated by our model are shown in Fig. 1.

## Computational evaluation of image quality

We employed two machine learning-based methods to evaluate the quality of a large number of generated images. First, we aimed to determine whether an image depicts a face or is considered a fail case, and second, we assessed whether characteristic features of the disorders are present in the images.

We generated 1000 random images for each class, including the unaffected faces class. While most generated images are high-quality portraits, some fail to convey meaningful content. We defined a fail-case as an image without any visible face. To estimate the number of fail-cases, we utilized RetinaFace, which was already used for image alignment. Since RetinaFace predicts facial features like eyes, nose, and mouth, its confidence can be considered a criterion for quality assessment; we considered an image a fail-case if the confidence of RetinaFace was below 99.9% (Supplemental Fig. 2). The percentage of fail-cases is below 10% for most disorders. Still, the proportion of fail-cases varies between disorders, such as 2.7% and 6.9% for Cornelia de Lange and Kabuki syndrome, respectively (Supplemental Fig. 3). Possible reasons for this could include lower training image quality, such as a low resolution or
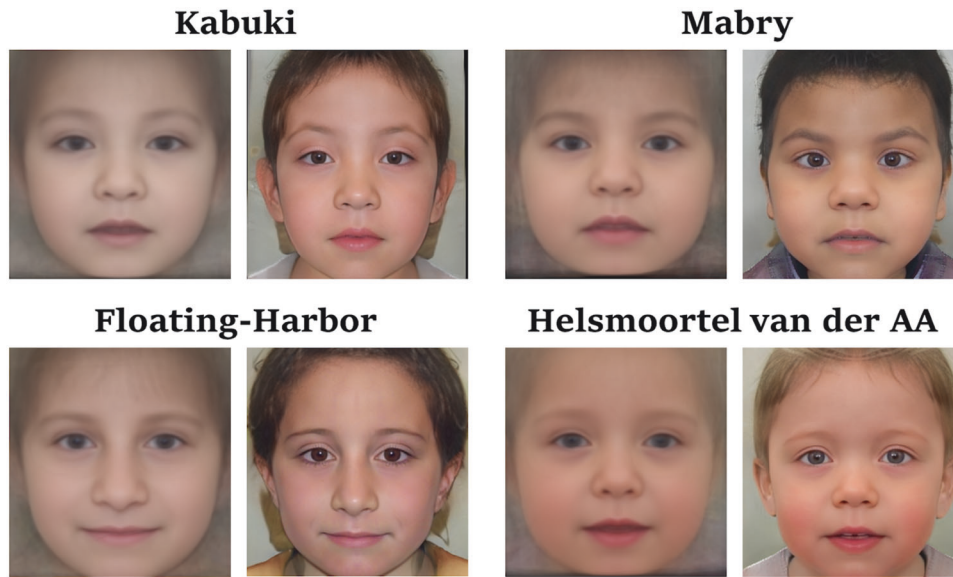
**Kabuki** **Mabry** **Floating-Harbor** **Helsmoortel van der AA**

**Fig. 3** **Comparison of ordinary image averages and latent averages generated for four disorders GestaltGAN was trained on.** Since both averaging techniques operate in essence on the same underlying data, there is a high similarity of image averages (left) and latent averages (right) for each condition. However, averaging in image space blurs fine structures, while latent averages appear more photorealistic.



**Fig. 4** **The survey was presented to human participants to assess their ability to recognize generated images, specific training images, and specific genetic conditions.** The lower section shows the expected result for each question due to random chance and what was observed. The closer the observed and expected values, the harder the question. 1) Participants could identify original images slightly more often than randomly expected. 2) Participants could not identify which individuals were used for training. 3) Participants could recognize the characteristic features in original and synthetic images with comparable precision. Color code: Original images are depicted in black, original images not part of the training set in yellow, and generated images in blue.

black-and-white images, or more unique facial features in certain disorders. Unique features pose a challenge for the model, as they are encountered less frequently in the training data.

To assess whether characteristic phenotypes of the disorders are represented in the generated images, we utilized Gestalt-Matcher. We tested whether the generated disorder was within its top-5 predictions. Overall, GestaltMatcher achieved a top-5 accuracy of 76.7% on the synthesized images. The top-5 accuracy rates for Cornelia de Lange syndrome and Williams-Beuren syndrome were above 90%. In comparison, the correct disorder was only listed in the top 5 differential diagnoses for synthetic images of Baraitser-Winter in 62.1% and for Nicolaides-Baraitser in 39.9% of the cases. These performances agree with the accuracy rates measured on real data, in which top accuracy rates also differ per disorder depending on their distinctiveness. Therefore, the results indicate that the generated images' characteristic features are indeed present.

**Assessment of image quality by human experts**
In addition to the computational techniques we used to assess image quality, we also developed an online survey with questions in three categories. We recruited 63 users of the GestaltMatcher database for the survey. These users represent dysmorphologists and other medical professionals working on genetic and other rare disorders. The experiment enabled us to assess the performance of humans in distinguishing 1) synthetic images from non-synthetic images, 2) re-identifying original data (images) that were used for training, and 3) identifying (diagnosing) the correct disorder. Each survey question also included a time-limit between 15 and 30 seconds, depending on the question, to prevent participants from scanning images for tiny artifacts, which can occur in synthetic images. In addition, a skip button allowed participants to skip a question and continue with the next question. In total we recorded 63 sessions, in which ten questions were asked in each of the three categories. Out of the 1860 answers, we excluded 106 skipped questions and 151 timeouts, resulting in 1603 answers for further evaluation. For all experiments, the images and choices were randomly sampled. The full experiment setup is visualized in Fig. 4.

In the first category, participants were presented with four distinct portraits, of which three were generated by GestaltGAN. At the same time, one, the original, was an image of an individual

from the training set. Participants were asked to identify the original. In 33.3% of the cases, participants were able to find the original, exceeding the expected chance value of 25%. The null-hypothesis, that generated images are indistinguishable from real images, had to be discarded (1.9 $10^{-5}$ < 0.05, binomial test). However, this was expected since generated images often contain artifacts that expose them as artificial, and still, in most cases, participants could not identify the original.

In the second category, participants were presented with a portrait representing a specific condition, averaged from the latent space of GestaltGAN. The participants were shown three original portraits of individuals with the same condition, only one of which had been used during the training. They were asked to identify the individual in the training set, in case the participant is unable to identify the individual used for training, privacy is considered to be protected. In 33.7% of cases, participants answered correctly, and the null hypothesis that GestaltGAN generates images that do not violate patient privacy (at least in the tested approach) did hold (0.889 > 0.05, binomial test).

In the third category, a synthetic portrait was shown to the participants, and they were asked to choose the correct disorder from four different options. If the synthetic images accurately represented the disorders, we hypothesized that experts should be able to identify the correct disorder at approximately the same rate as real images. Remarkably, in 48% of the cases, the experts could correctly diagnose the patient based on a real portrait, while their accuracy was 48.5% based on a generated portrait. The null hypothesis that generated images do not exhibit any characteristic features of the simulated disorder could be rejected (0.001 < 0.05, $\chi^2$ test). In fact, diagnostic yield on the generated images was even slightly higher, possibly due to an enrichment of characteristic features due to the GestaltMatcher loss function.

## DISCUSSION

In this study, we explored the application of Generative Adversarial Networks in generating photorealistic portraits for rare disorders that preserve the characteristic clinical features and patient privacy. We presented GestaltGAN, a modified StyleGAN architecture, and demonstrated in a series of experiments that synthesizing photorealistic faces of individuals with rare genetic conditions is possible despite limited training data. Through careful data preparation and augmentation, we were able to generate photorealistic portraits that accurately represent the facial features of a syndrome. Specifically, oversampling and our custom loss function enabled us to train the model to reproduce the characteristic features of disorders more accurately.

Our evaluation encompassed computational assessments of image quality and human evaluations through an online survey of medical professionals. First, the evaluation of all images demonstrated overall sound quality, with only a small percentage that had to be removed. The responses of medical professionals further indicated that synthesized portraits enabled them to identify the intended condition and exhibited features similar to the original disorders. Additionally, participants had difficulty recognizing the original image used for training, suggesting that GestaltGAN can be used to preserve patient privacy. Using the latent space, we presented the novel latent representations for conditions that average features in the latent space and appear much sharper than simple averages of the faces.

Our study has several important limitations. Different aspects were evaluated independently for different images. However, the tests would have to be met for one image simultaneously. For example, a synthesized image must be high-quality and distinct while preserving the patient's privacy. On the other hand, since an infinite number of synthetic images can be generated, it is not feasible to test privacy protection. An additional constraint was that we assessed a limited number of conditions and focused on a single generative method. While we do not necessarily anticipate extending the study in these ways would yield very different results, assessing our approach more broadly would be interesting. In the future, it could be interesting to use more detailed labels, such as additional age labels or individual HPO terms, instead of solely the condition in question. This could provide the user with more specific control over the generated faces.

In conclusion, we find the parallels of GANs to traditional medical education striking. In medicine, trainees learn according to the mantra "See one, Do one, Teach one". Similarly, by training on a relatively small number of cases, GestaltGAN achieved proficiency in generating accurate images of individuals with genetic conditions. The quality of the internalized knowledge was shown as the GAN discriminator, and experts could no longer reliably distinguish artificial and real images. Overall, this work highlights the potential of GANs in the medical field to artificially synthesize data while protecting patient privacy.

## DATA AVAILABILITY

All training data for GestaltGAN was extracted from GMDB. Photorealistic synthetic portraits of 20 disorders can be found at https://thispatientdoesnotexist.org.

## CODE AVAILABILITY

We publish our code on GitHub: https://github.com/kirchhoffaron/gestaltgan.

## REFERENCES

1. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.
2. Hsieh T-C, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. Gestalt-Matcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022. https://doi.org/10.1038/s41588-021-01010-x.
3. Hsieh T-C, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, et al. PEDIA: prioritization of exome data by image analysis. Genet Med. 2019;21:2807–14.
4. Schmidt A, Danyel M, Grundmann K, Brunet T, Klinkhammer H, Hsieh T-C, et al. Next-generation phenotyping integrated in a national framework for patients with ultra-rare disorders improves genetic diagnostics and yields new molecular findings. medRxiv. 2023. https://doi.org/10.1101/2023.04.19.23288824.
5. Saranya A, Subhashini R. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decis Analytics J. 2023;7:100230.
6. Bowles C, Chen L, Guerrero R, Bentley P, Gunn R, Hammers A, et al. GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks. arXiv [cs.CV]. 2018. http://arxiv.org/abs/1810.10863.
7. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019;6:1–48.
8. Duong D, Hu P, Tekendo-Ngongang C, Ledgister Hanchard S, Lui S, Solomon BD, et al. Neural Networks for Classification and Image Generation of Aging in Genetic Syndromes. Front Genet. 2022;13:864092.
9. Karras T, Laine S, Aila T A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv [cs.NE]. 2018. http://arxiv.org/abs/1812.04948.
10. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. arXiv [stat.ML]. 2014. http://arxiv.org/abs/1406.2661.
11. Mirza M, Osindero S. Conditional Generative Adversarial Nets. arXiv [cs.LG]. 2014. http://arxiv.org/abs/1411.1784.
12. Lesmann H, Moosa S, Pantel T, Rosnev S, Hustinx A, Javanmardi B, et al. GestaltMatcher Database - a FAIR database for medical imaging data of rare disorders. medRxiv 2023. https://doi.org/10.1101/2023.06.06.23288824.
13. Hustinx A, Hellmann F, Sümer Ö, Javanmardi B, André E, Krawitz P, et al. Improving Deep Facial Phenotyping for Ultra-rare Disorder Verification Using Model Ensembles. arXiv [cs.CV]. 2022. http://arxiv.org/abs/2211.06764.
14. Wang X, Xie L, Dong C, Shan Y. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* 2021: 1905–14.
15. Kang X, Yang T, Ouyang W, Ren P, Li L, Xie X. DDColor: Towards photo-realistic image colorization via dual decoders. *ICCV* 2022: 328–38.

382

16. Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: Single-stage Dense Face Localisation in the Wild. arXiv [cs.CV]. 2019. http://arxiv.org/abs/1905.00641.

17. Or-El R, Sengupta S, Fried O, Shechtman E, Kemelmacher-Shlizerman I. Lifespan Age Transformation Synthesis. In: *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp 739–55.

18. Karras T, Aittala M, Laine S, Härkönen E, Hellsten J, Lethinen J, et al. Alias-free generative adversarial networks. Adv Neural Inf Process Syst. 2021;34:852–63.

19. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25:60–64.

20. Xia W, Zhang Y, Yang Y, Xue J-H, Zhou B, Yang M-H. GAN Inversion: A Survey. IEEE Trans Pattern Anal Mach Intell. 2023;45:3121–38.

## AUTHOR CONTRIBUTIONS

All authors conceived the work, and played an important role in interpreting the results, drafted the manuscript, approved the final version, and agreed to be accountable for all aspects of the work. Aron Kirchhoff designed and implemented the architecture of GestaltGAN.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICAL APPROVAL

Ethical approval for the GestaltMatcher Database was granted by the IRB of the University Hospital Bonn.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41431-025-01787-z.

**Correspondence** and requests for materials should be addressed to Peter Krawitz.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.