

Synthetic Visual Data Generation and Analysis of Rosacea from Limited Data

Anwasha Mohanty, B.Tech, M.Tech

Supervised by Dr. Hossein Javidnia, Dr. Alistair Sutherland,
and Dr. Marija Bezbradica



A thesis presented for the degree of Doctor of Philosophy

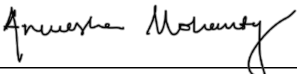
SCHOOL OF COMPUTING
DUBLIN CITY UNIVERSITY

January 2024

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Date: **09/01/2024**

Signature of the Candidate: 

ID No: 18211427

Acknowledgements

I wish to extend my deepest appreciation to everyone who has contributed markedly to the culmination of my PhD journey. Primarily and most importantly, I am sincerely grateful to my supervisor, Dr. Hossein Javidnia, for his outstanding mentorship and timely assistance whenever I encountered challenges. I especially thank Hossein for his insightful questions and for pushing me to deliver my best. A special thanks to Dr. Alistair Sutherland for his unwavering motivation, dating back to 2017 when I began applying for scholarships to pursue this doctoral journey in Dublin. I am sincerely grateful for his feedback on my drafts, and continued engagement with my progress even after retirement. I also extend my sincere gratitude to Dr. Marija Bezbradica for her unwavering support and valuable feedback on my drafts, even addressing my crucial concerns during her extended leave.

I am indebted to Prof. Frank Powell and the Charles Institute of Dermatology for generously providing early motivation and the Rosacea dataset essential for this research endeavour. I am immensely grateful to the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) for their support under Grant 18/CRT/6224. I would also like to thank Prof. Alan Smeaton for his encouragement, particularly during my transfer examinations and subsequently.

Angela Lally's incredible program management skills, which ensured seamless logistics, have been invaluable. I am equally grateful for her role as the go-to person for admin-related challenges, consistently providing solutions throughout this journey. I extend my heartfelt gratitude to Steven Carroll for organising the impactful writing retreat, which significantly kickstarted my thesis writing process. Many thanks to Tom Doyle for his prompt resolution of my complex hardware and IT-related issues, allowing me to stay solely focused on my experiments.

I am thankful to Dr. Jonathan Pearson and Dr. Dan Schofield for providing me with the invaluable opportunity to serve as a research intern with NHS England. I am also grateful to Dr. Alistair Sutherland and Dr. Graham Healy for providing me with the opportunity to instruct/teach several courses, allowing me to immerse myself in the classroom environment alongside this journey. I am also thankful to Prof. Rozenn Dahyot for her warm welcome into her team at Trinity College Dublin during the early stages of my PhD.

To my lab mates – Phuc, Mikhail, Bulat, Sidra, Eric, Hamza, Sanjay, Prashant, Carles, Yasser, and others – your camaraderie, lively discussions, and luncheons have not only enriched my journey but also made my day-to-day life in the lab easier, more interesting, and truly enjoyable, making the time seem to just fly by in such delightful company. I am also grateful for the early support of Fouad, Rehan, Artem, Rashmi, as my initial lab mates when I joined DCU. A special acknowledgment to my cohort mates, especially Ayushi, Shubhajit and Théo; our intermittent interactions and catch-ups evolved into valuable friendships over time. I am thankful to Sam,

Ruthanna, Math, Yongda, and Anish for being wonderful flatmates and for sharing delightful meals.

Asma's companionship, particularly her comforting meals during my highs and lows, holds a special place in my heart. I am thankful to Mark for recommending enriching books and to Jocasta for our memorable film-photography walks. My heartfelt thanks to my long-distance friends – Anisha, Chetna, Sonali, Ankita, Sonali, Faizan – for keeping our connections alive, and continuous support. I extend my deepest appreciation to Dexmont and Alireza, as senior colleagues, for their enduring support, encouragement, and for the delightful conversations and meals we have shared throughout this journey. I am grateful to Prof. Addapalli V.N. Krishna and Dr. Mausumi Goswami, who were esteemed faculty members during my Master's, for their continuous encouragement to undertake this PhD journey.

Immense gratitude to my parents, Ashok and Binodini, for their inspiration, encouragement, and unwavering love. My mother's constant endeavours to understand my journey have been heartwarming and strengthening. Conversations with my sister, Aparna, have been my respite, always cheerful and refreshing.

Lastly, I extend my deepest thanks to my best friend, Nadeem, for his enduring patience, unwavering support, and for joining in the celebration of every milestone and comforting in every setback of this journey. His companionship has been my steadfast rock, and his presence has been the anchor, keeping me grounded and propelling me toward achieving what I set out to accomplish.

I must express my appreciation for the cafes, bookstores, and film roll developers in Dublin; they have been indispensable to my well-being.

While it is challenging to name every individual, I extend my profound gratitude to everyone who has been a part of this journey. Every contribution, big or small, has been pivotal and instrumental in reaching this milestone. My heartfelt thanks go to each one for their individual and collective efforts and invaluable support.

Contents

1	Introduction	1
1.1	Motivation	5
1.1.1	Rosacea	6
1.1.2	Data Scarcity	9
1.2	Synthetic Visual Data	10
1.3	Usage and Benefits of Synthetic Visual Data	14
1.4	Challenges and Objectives	16
1.5	Hypothesis and Research Questions	18
1.6	Contributions	19
1.7	Thesis Structure	21
2	Background and Related Work	23
2.1	Digital Imaging in Dermatology	26
2.1.1	Motivation for Rosacea image analysis	28
2.2	Related work on Rosacea	30
2.3	Learning - from “Big Data” to “Small Data”	35
2.3.1	Big Data in computer-aided medical diagnosis	36
2.3.2	Importance of the ‘small datasets’ in medical diagnosis	37
2.4	Availability of Skin Disease Datasets and Challenges	43
2.5	A brief Overview of Skin Disease Analysis using Machine Learning and Computer Vision methods	45
2.5.1	Data Augmentation and Transfer Learning	47
2.5.2	Generating synthetic images using GANs	53
2.5.3	Meta-Learning and Few-Shot Classification	60
2.5.4	Federated Learning	66
2.5.5	3D face modelling	73
2.6	Discussion and outlook	77
2.7	Conclusion and Future scope	78
3	High-Fidelity Synthetic Rosacea Data Generation	80
3.1	Introduction	80
3.1.1	Importance of studying full-face images of <i>Rosacea</i>	81
3.2	Developments in Synthetic Face Generation	82
3.3	Methodology	87
3.3.1	StyleGAN2 with Adaptive Discriminator Augmentation	87
3.3.2	The impact of R_1 Regularization γ for 300 images	91
3.3.3	<i>Rosacea</i> Data	94
3.4	Experiments and Results	96

3.4.1	Truncation Trick	103
3.5	Qualitative Evaluation of Synthetic Images	109
3.6	Limitations and Discussion	112
3.7	Future work	115
3.8	Conclusion	116
3.9	Data and Code Availability Statement	117
4	3D Modelling of Rhinophyma	119
4.1	<i>Rhinophyma</i> Diagnosis	122
4.2	<i>Rhinophyma</i> Data	123
4.3	Towards 3D Face Modelling Approach	124
4.4	Methodology	125
4.4.1	3D Models Design Approach	125
4.4.2	<i>Rhinophyma</i> 3D Models	127
4.4.3	Rendering Set-up	128
4.5	Synthetic <i>Rhinophyma</i> images for Classification	131
4.5.1	Synthetic Data Preparation	132
4.5.2	Normal Nose for Classification task	134
4.5.3	Real-world <i>Rhinophyma</i> data preparation	134
4.5.4	Classification Models	136
4.5.5	Results and Discussion	137
4.5.6	Class Activation Maps	138
4.6	Limitations and Discussion	140
4.7	Utility of 3D Modelling and Synthetic Data Generation of <i>Rhinophyma</i>	142
4.8	Conclusion and Future Scope	143
4.9	Data and Code Availability Statement	144
5	Conclusion	146
5.1	Dissertation Overview by Revisiting Research Questions	146
5.1.1	Revisiting Research Question- 1	147
5.1.2	Revisiting Research Question- 2	151
5.1.3	Revisiting Research Question- 3	153
5.1.4	Revisiting Research Question- 4	153
5.1.5	Revisiting Research Question- 5	155
5.2	Limitations	157
5.3	Future Work	159

List of Figures

1.1	The illustration shows the full faces of Rosacea subtypes at the top, with their localized regions displayed at the bottom. Subtype 1 is characterized by persistent central facial erythema and frequent flushing with redness on the cheeks. Subtype 2 exhibits erythematous dome-shaped papules, some with surmounting pustulation in a centrofacial distribution (forehead, cheeks, and chin) against a background of persistent erythema (subtype1). Subtype 3 displays persistent facial swelling accompanied by hypertrophy of nasal tissue causing anatomical deformation, commonly known as Rhinophyma.	8
2.1	Clinical images vs. Dermoscopic images. The images used in this figure are taken from DermnetNZ [75].	28
2.2	An example of Meta-Learning set up in 1-shot-5-way classification.	62
2.3	Standard reference shape. This image is taken from the work by Lüthi et al.[152]	75
2.4	Generated face models using a Gaussian Covariance function with different standard deviations. This image is taken from the work by Lüthi et al.[152]. The model parameters include the scale factor s and the bandwidth σ , where s determines the variance (i.e., scale) of a deformation vector. Choosing a large σ leads to smooth, global deformations of the face, whereas a smaller σ yields more localized deformations.	76
3.1	Progress of synthetic face generation using various GAN models with the maximum volume of dataset available.	83
3.2	An example of blob-like artifacts in the generated images. This image is taken from Karras et al.[43]	85
3.3	The flow diagram of Stochastic Discriminator Augmentation [194], where G is the generator and D is the discriminator. The red boxes represent the 18 augmentation operations. The set of selected augmentation are controlled by the augmentation probability ' p ' and these augmentation can be visible to the discriminator D in the green box. The blue boxes represent the networks that get trained during the training process and yellow boxes represent the loss calculated after the training. In this set up, the non-saturating logistic loss is accommodated to calculate the final probability of the images being predicted as fake.	90

3.4	Kernel Inception Distance (KID) Progression for 10 Experiments Over Training Period. This graph tracks the KID metric's changes across different training steps for ten distinct experiments. Each coloured line corresponds to an experiment number as indexed in Table 3.1, allowing for a visual comparison of how each experimental setup impacts the KID over time.	104
3.5	Generated Faces from Diverse Training Strategies. This figure presents a visual representation of the synthetic faces generated across 10 different experiments as outlined in Table 3.1. Each sub-figure (a-j) corresponds to a unique combination of training setups, augmentation methods, freeze-D and regularization strengths γ detailed in the table. The images showcase the variation in quality and features of the generated faces. This figure, in conjunction with Table 3.1, provides a comprehensive comparison of how each methodology impacts the quality and realism of synthetic image generation.	105
3.6	Generated faces from the best KID value (3.1) of Exp 7 with the truncation $\psi=0.7$	107
3.7	Generated faces from the best KID value (2.5) of exp 10 with the truncation $\psi=0.7$	108
3.8	A 3D representation of Dermatologists opinion on synthetic images. .	109
3.9	Mean scores from the Dermatologists for generated images	110
3.10	Representation of mean scores (in %) for percentage of given images in the study by Dermatologists	112
3.11	Mean scores from the non-specialist participants for miscellaneous images	113
4.1	Illustration of 3D head models for both female and male, with severity level 0 (no deformations), representing typical facial features unaffected by <i>Rhinophyma</i>	126
4.2	An illustration of female and male head models with 6 nose variants at 3 levels of <i>Rhinophyma</i> severity.	129
4.3	An illustration of female and male 3D head models with added deformations that present patterns of <i>Rhinophyma</i> . Sub-figure (a) displays the female head model with a deformation applied to nose variant 1 at severity level 2.90. Similarly, sub-figure (b) presents the male head model with a deformation applied to nose variant 1 at the same severity level.	131
4.4	Generated images of the female (left) and male(right) images after rendering through the camera no. 5 (on the top), 6 (middle) and 7 (bottom).	133
4.5	Demonstration of 68-landmark points on the top subfigures illustrated in Fig.4.4	134
4.6	Example images of the female (left) and male(right) images through the camera no. 5 (on the top), 6 (middle) and 7 (bottom) after the Parsing technique has been applied.	135

4.7 Visualisation of feature maps produced by the classification models using GradCAM[295]. Images used in this visualization are taken from SD-260 [126]. The top two rows consist of images that are parsed according to the parsing technique discussed in Section 4.5.1. On the other hand, the bottom two rows feature manually cropped images with an emphasis on the nose region, as discussed in Section 4.5.3. 140

List of Tables

2.1	Number of dermatologists in six different countries. This table quantifies and compares the number of dermatologists available in six different countries, expressed per 100,000 population.	24
2.2	Comprehensive Summary of Research Studies on Rosacea and Associated Skin Conditions. This table encapsulates a detailed survey of research efforts undertaken to understand and classify Rosacea alongside various other skin conditions. It outlines each study by the author(s), publication year, the specific skin diseases addressed, dataset details including source and availability, volume of data used, the approaches taken to address the problem, methodologies applied, performance metrics attained, and deployment strategies.	38
2.3	Comprehensive catalog of accessible skin disease datasets. This table provides an extensive overview of publicly accessible datasets pertinent to skin disease research, including Rosacea. It enumerates each dataset, detailing its name, covered disease categories or specific conditions, imaging modality employed, total volume of data, number of classes featured, specific mention of Rosacea images if applicable, accessibility status, and originating country or region.	42
2.4	Comprehensive Summary of Research on Data Augmentation and Transfer Learning in Skin Disease Classification. This table delineates a curated list of studies that have leveraged both data augmentation and transfer learning techniques to enhance the accuracy and reliability of skin disease classification models. It details the authors, publication year, specific skin conditions targeted, types of data augmentation and transfer learning methods used, datasets employed, and the performance metrics achieved. The table further illustrates how the combination of data augmentation and transfer learning contributes to overcoming challenges such as data scarcity and model generalizability in dermatological imaging, reflecting the synergy between these two methods in advancing the field of medical diagnostics.	50

2.5	Detailed Analysis of Studies Utilizing Generative Adversarial Networks (GANs) in Skin Disease Classification. This table provides an in-depth overview of significant research where GANs have been applied to the field of dermatological imaging. It includes essential details such as the authors, publication year, specific skin diseases studied, names and sources of datasets used, dataset volumes both in total and per class, and the methodologies or benchmarks employed in each study. Additionally, the table highlights the best results achieved in terms of evaluation metrics, the number of samples generated, deployment strategies, and hardware used. This comprehensive summary offers insights into how GANs have been effectively utilized to generate synthetic data for skin disease classification. . . .	56
2.6	Comprehensive Review of Studies Utilizing Meta Learning and Few-Shot Classification in Skin Disease Diagnosis. This table collates a selection of significant research efforts that have applied meta-learning and few-shot classification strategies to tackle skin disease diagnosis challenges. It catalogs the studies by author, publication year, the specific skin diseases investigated, and the datasets utilized, including the names, sources, and volumes both in total and per class. Moreover, it details the methodologies and benchmarks used in each study along with the best results achieved, outlining the evaluation metrics, as well as deployment strategies and hardware requirements.	64
3.1	List of experiments and results detailing the training setups, augmentation methods, regularization strengths (denoted by γ), and the corresponding performance metrics. The table showcases different training strategies, including training from scratch and Transfer Learning (TL) from the Flickr-Faces-HQ (FFHQ) dataset, with varying degrees of augmentation choices and freeze-D technique. Performance is evaluated using the Kernel Inception Distance (KID) metric \downarrow , with the best KID score achieved noted for each experiment, along with the training step number at which this score was recorded. This comprehensive overview allows for a direct comparison of the impact of training methodologies and augmentation techniques on the quality of synthetic image generation.	100
3.2	Qualitative Evaluation: Dermatologists' Comments on Generated <i>Rosacea</i> Faces. This table presents a collection of feedback and insights from dermatologists regarding the synthetic <i>Rosacea</i> faces generated as part of this study. Each entry lists the comments provided by individual dermatologists, reflecting on the realism, potential utility, and overall impression of the generated facial images with <i>Rosacea</i> patterns. These comments are instrumental in understanding the clinical relevance and educational value of the synthetic images, as well as guiding future improvements and applications in dermatological training and diagnosis.	111
3.3	The FID values are calculated and compared with the top two experiments, selected based on having the lowest KID values.	114

4.1	<i>Rhinophyma</i> classification by el AZHARY et al. [278], score and range of severity by Wetzig et al. [279] scores (RHISI- Rhinophyma Severity Index) along with skin features as a design reference for 3D modelling.	125
4.2	The classification results. The scores calculated using various metrics on the real test set are highlighted in orange.	137
4.3	Confusion matrices for each classification model based on the Second Test Set. The Positive Class is “Normal Nose” and the Negative Class is “ <i>Rhinophyma</i> ”.	138

List of Publications

The following publications have been produced during the course of this research:

1. A. Mohanty, A. Sutherland, M. Bezbradica and H. Javidnia, “Skin Disease Analysis With Limited Data in Particular Rosacea: A Review and Recommended Framework,” in *IEEE Access*, vol. 10, pp. 39045-39068, 2022, doi:10.1109/ACCESS.2022.3165574.
2. A. Mohanty, A. Sutherland, M. Bezbradica and H. Javidnia, “Towards Synthetic Generation of Clinical Rosacea Images with GAN Models,” 2022 33rd Irish Signals and Systems Conference (ISSC), Cork, Ireland, 2022, IEEE pp. 1-5, doi: 10.1109/ISSC55427.2022.9826207. (Best Poster Presentation Award)
3. Mohanty, A., Sutherland, A., Bezbradica, M. and Javidnia, H., 2023. “High Fidelity Synthetic Face Generation for Rosacea Skin Condition from Limited Data”. arXiv preprint arXiv:2303.04839. (Accepted for publication in *Machine Learning in Electronic and Biomedical Engineering, Volume II, Electronics in 2024*)
4. Mohanty, A., Sutherland, A., Bezbradica, M. and Javidnia, H., 2023. Rhi3DGen: Analyzing Rhinophyma using 3D face models and synthetic data. *Intelligence-Based Medicine*, Elsevier p.100124.

Synthetic Visual Data Generation and Analysis of Rosacea from Limited Data

Anwasha Mohanty

Abstract

Skin diseases, encompassing approximately one-third of global human ailments, remain the fourth leading cause of global disease burden. Specifically, Rosacea, despite its significant prevalence, suffers from a marked scarcity in clinical visual data. Data scarcity, impedes the effective utilisation of deep learning models in computer-aided skin disease diagnosis, especially for conditions often overlooked in clinical visual/image data acquisition.

This study meticulously addresses Rosacea’s data scarcity issue. An exhaustive literature survey spotlighted Synthetic Visual Data Generation as a potential solution to this data deficit. The central aim is to innovatively acquire and process data, mitigating the ramifications of visual data inadequacy by producing high fidelity synthetic visual data across three Rosacea subtypes.

To the best of our knowledge, this constitutes the first attempt to employ Generative Adversarial Networks (GANs) with such a limited dataset of 300 images—a scenario in which GAN models typically struggle to converge. However, leveraging the theoretical principles of GANs enabled successful model convergence and the generation of high-fidelity Rosacea images using a variant of StyleGAN2. Furthermore, we have, for the first time, innovatively employed the concept of 3D Parametric Modelling and computer graphics, facilitating the construction of 3D head models for subtype-3 using only 268 images.

The application of these techniques successfully generated synthetic data for Rosacea Subtypes-1, 2, and 3. For subtype-1 and 2, Board-certified expert dermatologists and lay participants validated the synthesised images. For Subtype-3, the efficacy of synthetic data was further corroborated by classification models, emphasising the viability of synthetic data when juxtaposed with real-world images. Grad-CAM visualisations provided additional validation of these models’ robustness. The resultant high-fidelity datasets for Rosacea Subtypes-1, 2, and 3, now publicly accessible, affirm the proficiency of synthetic image generation in addressing the challenges of data scarcity inherent to conditions like Rosacea.

Chapter 1

Introduction

Artificial Intelligence (AI), frequently proclaimed as the digital era's steam engine for its unparalleled potential to reshape every sector and stimulate significant economic growth, continues to leave a mark. While AI encompasses a wide array of technologies, it is Deep Learning [1], a subfield of Machine Learning [2] that is particularly revolutionary for its thorough development in modelling and comprehending intricate patterns. Deep Learning, by addressing numerous tasks in the fields of natural language, speech, and computer vision, stands as a foundational pillar and key driver of the ongoing advancements in AI. Nevertheless, the power of today's Deep Learning algorithms is contingent on the availability of extensive datasets that have been used for their development. Therefore, the production of data and the advancement of deep learning models are fundamentally interconnected.

Particularly in the field of computer vision, the availability of large datasets has been a significant driving force behind the success of deep learning. ImageNet [3], one of the largest publicly available vision datasets, was introduced in 2009 initially containing around 3.2 million images. Over the years, it has significantly expanded. As of the latest figures on the ImageNet homepage, the dataset has grown to contain more than 14 million images, encompassing over 21 thousand synsets (groups/classes). The initial 3.2 million images provided a substantial base, and the continuous growth to over 14 million images demonstrates its ever-increasing scale. ImageNet contains both everyday and obscure objects and specific items

including various species of animals (like ‘cat’, ‘elephant’), types of vehicles (like ‘car’, ‘airplane’), everyday objects (like ‘chair’, ‘table’), and even abstract concepts or events (like ‘birthday party’, ‘flash mob’) etc. Thus, the ImageNet dataset has served as the basis for approximately 55,000 research articles (using citations as a metric) and has been a major contributor to the AI revolution in the field of computer vision.

While ImageNet alone does not solve all global problems, having access to such an expansive dataset has indeed ignited a wealth of ideas for addressing various vision tasks, not just in machine learning but also in other fields dealing with visual data e.g. medicine, automobiles, astronomy etc. The medical field is one of the many fields that extensively relies on visual data in day-to-day decision making [4]. The spectrum of visual data in this field encompasses areas such as medical imaging, pathology, ophthalmology, surgical planning and navigation, and dermatology, all of which greatly influence diagnostic decisions.

With the advent of large labelled datasets such as ImageNet [3], that involve millions of images, complex image recognition and classification tasks have become feasible. Therefore, there is a general expectation to improve medical diagnosis with the aid of advanced deep learning and computer vision algorithms. Nevertheless, medical datasets are usually smaller, ranging in the hundreds or thousands [5]. While it may seem like the solution is simply to gather more data within the medical vision field, unfortunately, the process is not as straightforward as collecting images of everyday objects. The main obstacle in acquiring large datasets within the medical or clinical domain arises from the complex nature of the data acquisition itself. This process is heavily influenced by an array of factors such as the type of disease, geographic location, the type and settings of imaging equipment, time constraints, patient privacy concerns, and copyright restrictions of the institution, to name just a few. As a result, from a statistical perspective, various medical image datasets are susceptible to exhibiting a long-tailed distribution i.e. many disease classes may have only a small number of examples. The long-tailed distribution observed in the

medical vision field can pose difficulties for machine learning algorithms and models. This is because these models might be biased towards the disease categories, e.g. cancer, where there is a large amount of data available covering all the possible disease classes within the category and leading to suboptimal performance on the disease categories with lower occurrences of patterns.

Similar to the challenges in data acquisition, deep learning algorithms based on vision data also present notable limitations [6, 7]. These include the requirement for extensive volumes of labelled data, sensitivity to long-tailed distribution of datasets and out-of-distribution data, and a lack of interpretability. Further limitations encompass a lack of robustness and an intensive demand on resources - including time, space, and cost [8, 9, 10]. These factors contribute to the ongoing challenges in effectively deploying deep learning algorithms in medical applications.

While the intersection of medical vision data, deep learning, and computer vision algorithms has led to significant advancements in disease detection and diagnosis, these achievements are often concentrated on specific, high-profile diseases. These diseases, which I refer to as ‘prominent diseases’, are generally well-known, have severe consequences, and have benefited from extensive data accumulated over years by leading institutions and medical professionals. In contrast, there remains a vast array of disease categories, known as ‘tail classes’, that lack this level of attention and data acquisition. Often, these conditions are only addressed by a handful of organizations and healthcare professionals, which results in a starkly lower amount of data compared to the more popular diseases. Despite this, these ‘tail classes’ diseases are equally deserving of attention and diligent diagnosis care. Therefore, finding ways to improve the data situation for these less-studied conditions remains a significant challenge in the field.

A viable method to start addressing the long-tailed and out-of-distribution nature of ‘tail-class/rare class’ of disease categories in the medical imaging domain is to focus on one disease class at a time. This strategy facilitates the development of specialized models, optimized to make accurate predictions for individual disease

classes. By concentrating on specific diseases, we can identify their unique characteristics and tailor the data or models to better comprehend their traits more accurately.

Obtaining images in substantial quantities can prove to be a daunting task, especially for those that represent rare and critical events. The complication heightens when the objective is to train a detector for an entirely new category. The assembly and annotation of an exhaustive training set, inclusive of that category and its intra-class variations, could be resource-intensive. A similar obstacle arises in the absence of an existing dataset suitable for the intended task, leading to the necessity of creating and labelling a new dataset from the beginning. In fact, the majority of AI and deep learning projects, the tasks of data preparation and engineering consume over 80% of the time [11, 12], with this proportion escalating further as tasks become increasingly specific. However, the advent of synthetic data offers a potential paradigm shift in this landscape. Synthetic data is artificially generated data that can replicate the statistical properties of real-world data, allowing for the creation of large, diverse, and detailed datasets without the need for extensive collection and labelling efforts. By using techniques like computer graphics, generative models, and simulation, synthetic data can be produced to match specific requirements, including rare or critical event representation, thereby potentially reducing the time and resources needed for data preparation and engineering significantly. While synthetic data can indeed streamline the process, the extent of its impact is contingent upon the quality, realism, and relevance of the generated data to the task at hand. Effective use of synthetic data also requires validation against real-world data to ensure accuracy and performance, which can introduce new challenges and considerations. Thus, while synthetic data holds promise for reducing the proportion of time spent on data tasks, the actual impact would depend on various factors including the task complexity, data generation methods, and the necessity for validation and refinement.

Even when focusing on one disease class at a time, addressing the problem of data

scarcity through further data collection may not be practical given the time it takes for such collection. Indeed, data acquisition in the medical field for the creation of datasets suitable for computer vision applications is a time-consuming and expensive process. It typically involves seeking patient approval, collecting data, annotating or labelling the data, and adhering to ethical guidelines, among other tasks, all of which vary depending on the nature of the data. For instance, collecting natural images like those in ImageNet took about three years of initial data collection and annotation by non-expert annotators. In contrast, medical data needs to be annotated by doctors, a task made difficult by their often busy schedules. While data collection is certainly not discouraged, the focus of this discussion is on exploring what can be done with the data that is already available, particularly for datasets that are underrepresented in both the medical and vision research communities, even for common diseases.

1.1 Motivation

A vast number of diseases, despite their common occurrence, are not sufficiently represented in available medical vision datasets. This reality presents a clear and compelling motivation for this research study. With the rapidly increasing prevalence of numerous health conditions globally, it becomes paramount to ensure that medical and vision research caters to a broader spectrum of diseases, rather than being limited to only those that are currently well-documented in terms of data collection.

Among these underrepresented diseases, skin diseases, in particular, pose a significant challenge in the field of medical diagnosis due to their intricate observational and analytical requirements. Skin diseases are ubiquitous and have an enormous impact on people's lives, from the general population to all ages and ethnicities worldwide. Diagnosing skin diseases necessitates extensive experience and expertise, as the process often involves initial visual screening, subsequent dermoscopic analysis, biopsy, and histopathological evaluation. This conventional diagnostic process is both time-intensive and costly. A solution to this could be computer-aided

diagnosis. However, the relative scarcity of visual data for many of these diseases, particularly in the context of computer vision applications, makes them a challenging yet crucial area for research.

These complications are further exacerbated in cases of chronic skin conditions, which typically demand continuous engagement with a dermatologist over a patient's lifetime. Early detection of such chronic skin conditions can facilitate timely intervention, potentially averting more severe complications. Yet, the wise saying '*prevention is better than cure*' is often unattainable for most patients due to the extensive wait times for medical appointments, rendering early-stage diagnosis challenging [13]. An in-depth exploration of the related statistics will be discussed in chapter 2 of this thesis.

1.1.1 Rosacea

Rosacea is a chronic facial skin condition characterized by cyclical periods of remission and relapse [14]. It is also identified as a cutaneous vascular disorder [15]. This skin condition is particularly prevalent among individuals hailing from northern countries with fair or Celtic complexions [16]. According to a study published by the British Journal of Dermatology, there are nearly 415 million people affected by Rosacea worldwide [17]. Rosacea is typically marked by symptoms such as facial flushing and redness, inflammatory papules and pustules, telangiectasias, and facial edema. Notably, the severity of these symptoms exhibits considerable variation among different individuals[18].

In the realm of medical diagnostics, Rosacea is segregated into four distinct subtypes: Subtype 1 (Erythematotelangiectatic Rosacea aka ETR), Subtype 2 (Papulopustular Rosacea aka PPR), Subtype 3 (Phymatous Rosacea aka Rhinophyma), and Subtype 4 (Ocular Rosacea)[19]. The diagnosis of each subtype hinges upon the severity of the condition, classified as mild, moderate, or severe [14, 20]. This classification enables a more nuanced understanding of the disease and facilitates the formulation of a targeted treatment approach. This study is focused on the

Subtype 1 (ETR), Subtype 2 (PPR) and Subtype 3 (Rhinophyma). These three subtypes are illustrated in Fig.1.1.

Erythematotelangiectatic Rosacea (ETR) and Papulopustular Rosacea (PPR) are distinctive subtypes of Rosacea that, nevertheless, share a range of visual manifestations, a commonality that arises from their shared classification within the Rosacea family [19]. Each subtype tends to display persistent facial erythema, localized particularly to areas such as the cheeks, nose, forehead, and chin. Both ETR and PPR can also lead to the development of visible, small blood vessels, known as telangiectasias, on the face. These features are less common in the transient redness experienced by healthy individuals. The persistent nature of redness, specific facial patterns, visible blood vessels and associated symptoms such as burning or stinging sensations are key differentiators. Transient episodes of flushing, a symptom defined by a brief period of heightened redness in the skin, is a phenomenon present in both ETR and PPR. Furthermore, these conditions share a progressive nature, denoting a gradual worsening of symptoms over time if left untreated. There is a noted tendency for ETR, when deteriorating, to evolve into PPR, marking a transition from one subtype to another. Given the severity and progression of these conditions, early diagnosis, proper management, and appropriate medication are of paramount importance. In the absence of such measures, these subtypes may worsen rapidly, developing into chronic conditions that persist over extended periods and become resistant to complete resolution. Thus, the early management of these conditions is critical in preventing their transition to a chronic and potentially incurable state. It is crucial to understand the Erythematotelangiectatic Rosacea, is characterized not only by persistent central facial erythema but also by flushing and visible blood vessels, which are less common in transient redness experienced by healthy individuals. The persistent nature of redness, specific facial patterns, and associated symptoms like burning or stinging are key differentiators.

Rhinophyma, a subtype of Rosacea, is less prevalent compared to its counterparts, yet it manifests with drastic physical alterations to the skin [14, 19]. These

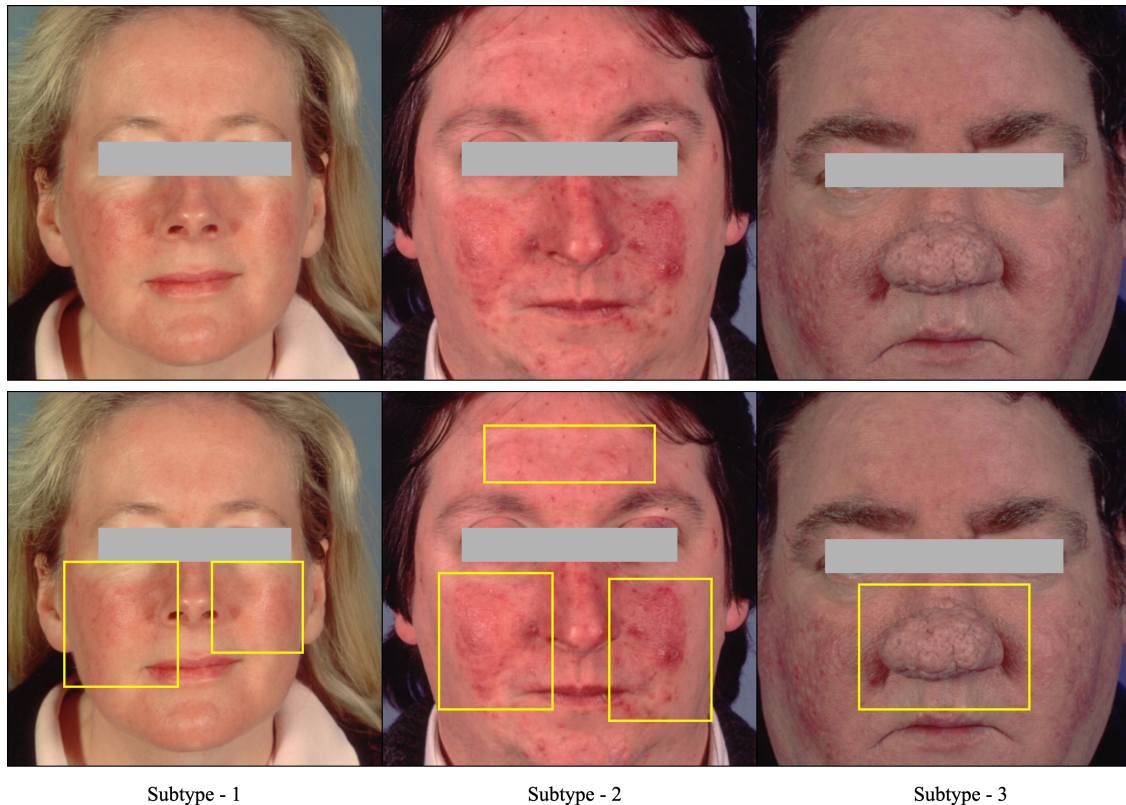


Figure 1.1: The illustration shows the full faces of Rosacea subtypes at the top, with their localized regions displayed at the bottom. Subtype 1 is characterized by persistent central facial erythema and frequent flushing with redness on the cheeks. Subtype 2 exhibits erythematous dome-shaped papules, some with surmounting pustulation in a centofacial distribution (forehead, cheeks, and chin) against a background of persistent erythema (subtype1). Subtype 3 displays persistent facial swelling accompanied by hypertrophy of nasal tissue causing anatomical deformation, commonly known as Rhinophyma.

alterations are typically characterized by skin thickening and enlargement, predominantly around the nasal area. Addressing the symptoms often involves medicinal interventions, while in severe instances, surgical or laser procedures may be necessitated to alleviate skin thickening and ameliorate the pronounced anatomical transformations [21]. It is evident that the efficacious management of symptoms can be achieved more effectively with timely detection and treatment. The early diagnosis not only helps in mitigating the symptoms but could also decelerate the progression of the condition.

Despite its widespread prevalence, Rosacea frequently remains underdiagnosed. Characteristic signs of Rosacea, such as facial redness and flushing, often result in misdiagnosis as seborrheic dermatitis, despite the two conditions being unrelated

[22]. Further, Rosacea is frequently mistaken for other dermatological conditions, including psoriasis, lupus, acne, and eczema [22, 23], thereby highlighting the challenges faced in its accurate and timely diagnosis. The reasons behind the continued growth of Rosacea cases in recent years are: lack of awareness in dermatologists, misdiagnosis, and the cost of treatment etc. [23]. In a recent article published by US News Health, “Rosacea is often misdiagnosed, and many do not seek treatment because they do not realize it is Rosacea,” says Jeffrey Fromowitz, MD, FAAD, a dermatologist based in Boca Raton, Florida [24]. Therefore, it is evident that Rosacea is often misdiagnosed due to factors such as insufficient knowledge about the condition among both dermatologists and patients. Given that Rosacea is primarily diagnosed visually in its early stages, there is undoubtedly a scarcity of available data for study and increased awareness. Details about Rosacea statistics – including global demographics of the disease, availability of expert dermatologists, average waiting times to consult a dermatologist, instances of misdiagnosis, the extent of publicly available datasets for research, the existing computer vision and deep learning based research conducted on Rosacea images – are thoroughly discussed in chapter 2 of this thesis.

1.1.2 Data Scarcity

In the previous subsection, we discussed the challenges related to acquiring medical vision data and the underrepresentation of certain diseases in the computer vision community. Given these circumstances, the amount of data available for underrepresented conditions, such as Rosacea, is significantly low, with available datasets often only numbering in the few hundreds. Therefore, it would be highly desirable to enhance learning strategies with the limited amount of available data, aiming to address the overarching challenge of data scarcity in the medical vision field. In this study, strategies are investigated to optimize the use of existing methods for learning from sparse data, with an aim to address the pervasive issue of data scarcity of the vision data in the medical domain. The main motivation to deal with the limited data in this research is the limited availability of datasets for Rosacea.

This study primarily addresses the issue of limited data by employing Rosacea as a representative case study. We have conducted a thorough literature review exploring numerous methodologies, focusing on those best suited to handle limited data in the context of deep learning. More specifically, this research explores the application of Generative Modeling techniques tailored to Synthetic Data Generation [12] for 3 subtypes of Rosacea skin condition. This comprehensive approach serves to increase the available visual data for Rosacea and enhances the capacity for more robust and reliable computer-aided diagnosis.

1.2 Synthetic Visual Data

Synthetic data refers to a category of data that is created artificially via computer algorithms and simulations, setting it apart from real data that is collected directly from the world around us. Synthetic visual data has played a pivotal role in the development and evolution of computer vision related tasks. Various methods exist to generate an array of synthetic data [12]. These methods include Reconstruction Techniques, Physics-based Simulators, 3D Modelling, Graphics-based Techniques and Deep Generative Models. Synthetic data generation and utilization are garnering significant attention, particularly due to its ability to generate data that closely mirrors the realism of actual data. This convincingly realistic synthetic data has demonstrated immense usefulness in numerous studies. The proliferation of synthetic data parallels the evolution of Deep Learning methodologies. However, the exploration into synthetic visual data predates the contemporary surge of deep learning. These foundational investigations have undeniably informed and influenced subsequent advancements.

In early computer vision history (1960s-70s), synthetic data was crucial for problems like line labelling. David Huffman’s classic research on “Impossible Objects as Nonsense Sentences” [25] presented the challenge of identifying concave and convex edges from images of embedded parallelepipeds. One image depicted all edges, including invisible ones, while another showed only the visible. The task was to

label such visible edges with ”+” for convex and ”-” for concave. Similarly, Maxwell Clowes, in “On Seeing Things” [26], introduced a related problem, discerning polyhedron shapes based on projected edges on a plane. Clowes focused on categorizing corners based on the type of their constituent edges. Intriguingly, both abstracted the problem of constructing these graphs, analogous to ‘edge detection’ in modern computer vision, and operated under the assumption of a pre-constructed line graph. However, this abstraction led to missing information in real images, pointing to a deficiency in edge detection. This highlighted the importance of synthetic data in their work to derive more accurate interpretations of real-world images. The necessity was addressed using 2D line drawings corresponding to feasible 3D objects. Notably, the algorithms in both papers did not require training sets; instead, they were tested solely on artificially produced line drawings, marking some of the earliest and simplest uses of synthetic data/images in computer vision.

Robotics, emerging shortly after artificial intelligence, took on the complex challenge of creating physical entities for real-world operations. Early in AI’s period in the 1970s, robotics became a central focus, leading to creations like “Stanford Cart” [27]. The Cart was equipped with a vision system, utilising an onboard TV. It was navigated using a ‘super-stereo’ vision algorithm. Designing such technology was costly, and validating their ideas in the real world was often entirely impossible. Before the 1990s, computer vision and robotics largely relied on hardcoded algorithms. Real-world validations were cumbersome and often unrealistic. The value of computer simulations was soon realized, given their efficiency and the capability to eliminate real-world noise and errors. This understanding initiated the “*simulate first, build second*” approach in robotics - a philosophy that now extends beyond robotics and has become regularly adapted in many other domains within AI.

Given this philosophy in robotics, in the 1980s - 1990s, with the rise of computer graphics, researchers started to understand the potential of synthetic data. The fundamental idea was simple; if real-world data was hard to acquire, why not generate artificial data. Thus, synthetic images began to be used as data points for

various vision tasks. In the 1980s, the use of computer-generated images primarily focused on fields like entertainment, simulations, and gaming. The value of these images and purposes quickly became noteworthy. Computer graphics and 3D modelling played an important role in generating synthetic data through methods like statistical shape modeling-analysis, procedural generation, voxel-based modeling, physics-based rendering methods were common techniques used for a broad range of real-world replications and further applications.

In the 1990s, building upon this trajectory in computational modelling, one technique that gained prominence, especially for medical imaging connected with computer vision, is Statistical Shape Modelling (SSMs) [28, 29]. It became particularly vital in areas that require a detailed understanding of shapes and their variations, such as medical imaging [30]. This methodology evolved into Active Appearance Models [29], which subsequently gave rise to 3D Morphable Models [31]. During the 1990s, these 3D models gained popularity for designing facial shape variations and were extensively used for human face modelling. The 3D Morphable Face Model is a generative model that represents both face shape and appearance. Designing these models often involves intricate steps and presents numerous challenges. Although these 3D Morphable Face Models were widely used for various applications, they were not adapted for any medical conditions on the face anatomy [32].

The 2010s marked a surge in the popularity and efficacy of deep learning techniques. Particularly, models based on neural networks introduced in early 1990s, like convolutional neural networks (CNNs) [33], demonstrated exceptional accuracy in image recognition. However, their optimal performance hinged on access to massive training datasets. In situations where gathering real-world data posed challenges or ethical concerns—such as medical imaging or surveillance—synthetic data emerged as a solution. Additionally, synthetic images enabled to be curated diverse datasets, thereby mitigating issues related to bias or under-representation.

While the use of synthetic data can address challenges related to time, cost, and efficiency of cutting-edge algorithms, it also poses certain issues. Reflecting on

earlier applications in robotics and self-driving vehicles, one of the major challenges was the domain gap between synthetic and real-world data. The question arises: how can we ensure that algorithms trained on synthetic data function effectively in real-world settings? To bridge this gap, the concepts of Transfer Learning [34, 35, 36] and Data Augmentation [37, 38] have become popular. Through these methods, researchers have devised techniques to transition models from synthetic to real-world domains. This approach harnesses the advantages of synthetic data without undermining performance in real-life scenarios.

Building on this momentum, in the 2010s, the introduction of Variational AutoEncoders (VAEs) [39] and Generative Adversarial Networks (GANs) [40] as Deep Generative Models marked a significant advancement in synthetic image generation. While VAEs excelled in faster generation and diversity, GANs could produce photorealistic high-resolution images that surpassed those of VAEs, blurring the lines between synthetic and real images. These generated images found various applications, from art and face generation to the creation of training datasets.

Over the last five years, GANs[40] have gained prominence across a multitude of applications [41, 42, 43, 44, 45, 46, 47]. Likewise, another commonly adopted approach for synthetic data generation involves the use of graphics-based simulations, such as 3D modeling, which have found extensive use in fields like object recognition [48, 49, 50, 51], scene understanding [52, 53, 54, 55, 56], and other applications [57, 58, 59, 60, 61].

Despite the broad application of synthetic data generation, only a handful of studies [61, 62] have ventured into generating full-face synthetics using computer graphics, where aspects such as identity, expression, texture, hair, clothing, and background environment were modeled. Nevertheless, the modeling of skin diseases has not been explored, largely due to the intricate nature and complexity of representing skin conditions at a granular level.

In this research, the generation of synthetic data for Rosacea subtypes is achieved using two approaches such as Deep Generative Modeling like GANs, 3D Model-

ing techniques like 3D Parametric models, and Computer Graphics, particularly in limited-data settings.

1.3 Usage and Benefits of Synthetic Visual Data

While real data usually provides the most insightful observations, it can often be costly, unevenly distributed, biased, inaccessible, or restricted due to privacy reasons. Synthetic data can effectively be used by offering more thoroughly annotated data for developing precise, adaptable models that frequently helps to overcome the shortcomings of the real data. Synthetic data can be employed for testing new ideas when real-world data is unavailable or when existing real-world data is skewed. Moreover, synthetic data can be used to enhance small datasets that might be currently overlooked. In cases where real data is unusable, non-shareable, or immovable, synthetic data presents a viable option. Thus, synthetic data plays a significant role in enabling further advancements.

Nowadays, several companies such as Microsoft, NVIDIA, Google, Meta, and to name but a few have been placing greatest importance on the use of synthetic data for training models. Moreover, the trend in industry solutions is increasingly leaning towards synthetic data generation, as dealing with real-world data can be messy and time-consuming. Additionally, to support the labelling process, numerous companies have begun to offer services related to synthetic medical data annotation and labelling. These services contribute to accelerating the medical decision-making process by leveraging the combined expertise of medical professionals and computer scientists within their teams.

Utilizing synthetic data for the training of medical/clinical visual data-driven decision-making models can potentially circumvent the substantial costs, restrictions, risks, and time associated with acquiring and labelling large quantities of real data. Given the medical system's constraints in many countries, obtaining visual data for various common disease categories can often be impossible. The concept of synthetic data generation paves the way for creating extensive examples of long-tail,

rare, high-risk, and safety-critical scenarios. However, it is essential to acknowledge that the quality and diversity of synthetic images are contingent upon the variety and representativeness of the source data. While synthetic data techniques can significantly augment a dataset, there are practical limits to the variability and realism that can be achieved, particularly when starting with a small set of real images. For instance, generating 1 million high-quality, diverse rosacea images from a few hundred real ones would face challenges in ensuring each synthetic image is unique, accurately represents the condition, and maintains high quality. The effectiveness of augmentation depends on factors like the algorithm's sophistication, the quality of the original dataset, and the specific requirements of the task at hand. Therefore, while synthetic data generation allows for substantial expansion beyond the original dataset size, it is more accurate to view it as a way to enhance and diversify data within certain bounds. Whether in a randomized class conditioning process such as Deep Generative Models or controllable such as a 3D modeling environment, this method enables the generation of even the rarest events with the same ease as more common ones. Furthermore, an extensive number of perfectly labelled data points can be generated quickly compared to the manual data gathering process. Once the data is generated, it can further be verified by medical professionals, human labellers, or through testing against a small set of available real-world data for the specific medical condition. Additionally, it would be beneficial to conduct a detailed analysis of the performance of models trained with various synthetic data. This involves experimenting with different characteristics of synthetic images, such as variations in lighting, angles, and levels of detail of specific disease patterns, as well as the types of synthetic augmentations used (e.g., geometric transformations, texture changes). By evaluating the model's accuracy, recall, precision and other relevant metrics in scenarios using different sets of synthetic images, we can identify patterns and characteristics that lead to improved or diminished performance. This empirical approach will provide valuable insights into the effective use of synthetic data in enhancing machine learning models, particularly in the domain of clinical

image analysis such as diagnosing Rosacea.

1.4 Challenges and Objectives

Diagnosing skin diseases using medical/clinical images with computer-aided decision making, especially in resource-limited conditions, inherently presents significant challenges. Addressing data scarcity in the medical field poses a formidable task, especially when trying to generate synthetic data using generative modeling approaches with only a handful of input data. Additionally, the usefulness of the generated synthetic data can vary, potentially rendering it advantageous or disadvantageous. The **primary anticipated challenges** in this research study include:

- Predominantly, advancements in computer vision techniques and deep learning algorithms are fueled by the availability of large datasets, leaving techniques that leverage small amounts of visual data less explored. This scenario prompts the question of which potential methodologies, models, and techniques could be effectively employed when only limited data is accessible.
- GANs typically achieve more stable convergence when provided with extensive input datasets. While numerous studies have demonstrated GANs' ability to process large datasets (in thousands) and achieve stable convergence, their capacity to converge with a more limited amount of data (specifically in the hundreds) is not as extensively documented. Although the foundation for this work on GANs stems from state-of-the-art research, the targeted exploration of their efficiency with such minimal data sets represents a niche yet to be thoroughly investigated. The primary challenge is to identify a GAN model configuration optimized for smaller datasets. While the model used in the research was adapted from existing work, the first contribution in this thesis will be to investigate its adaptation and efficiency in scenarios with limited image availability.

- A critical evaluation criterion for the study is the realism of the generated Rosacea images. It has not been determined up to now whether these synthesized images will meet the rigorous standards set by medical professionals in the field in terms of fidelity.
- The process of 3D modeling and synthetic data generation for a Rhinophyma nose from a limited collection of two-dimensional images presents an intriguing research problem. Exploring potential strategies and techniques to effectively achieve this transformation is a significant aspect of this study. A key question is whether it is feasible to design fine-grained deformations of Rhinophyma using 3D modeling techniques, a prospect that is as challenging as it is risky, given the necessity for subsequent design validation.
- The role of synthetic data generated through a 3D modeling approach in enhancing the performance of classification models is a subject of active research. The final challenge of this study is to investigate whether synthetic data generated in this study can improve classification models' ability to learn disease features more effectively by validating against a small number of real-world data instances.

The objectives of this research are the specific goals we aim to achieve to address the identified challenges in synthetic data generation for medical applications, particularly for dermatology cases. These objectives represent concrete steps we plan to undertake to overcome these challenges and advance the field of medical visual data in the computer vision research community. By achieving these objectives, we aim to contribute to the development and application of synthetic data in medical decision-making processes, particularly for conditions like Rosacea where the availability of visual data is limited. The main **objectives** of this research are:

- The primary objective of this research study is to address the crucial challenge posed by limited datasets for research. This entails identifying and investigating deep learning methodologies that have not been extensively studied or

surveyed for small data scenarios.

- A significant approach involves the use of Generative Adversarial Networks (GANs) to synthesize facial images exhibiting Rosacea characteristics from just 300 images. The main objective here is to find the optimal GAN configuration that allows convergence with this limited dataset. After synthetic image generation, it is crucial to validate these images with the help of domain experts. A qualitative method of verification would lend credibility to the generated images and ensure their suitability for further research.
- A key aspect of the proposed research is the conceptualization and realization of a Rhinophyma nose using a 3D modeling approach. This strategy facilitates the incorporation of fine-grained details characteristic of Rhinophyma, making the model highly representative of actual manifestations of the condition.
- The final objective is the verification of classification using synthetic images generated by the 3D modeling technique. This serves to investigate the potential of these artificially created visual data in enhancing the performance and benefits of the synthetic data under consideration.

1.5 Hypothesis and Research Questions

We hypothesize that certain GANs and 3D modeling techniques can effectively generate high-fidelity synthetic medical images from limited datasets of skin conditions like Rosacea and its sub-category Rhinophyma.

According to above hypothesis, the following **research questions** are formulated to guide the hypothesis:

RQ1 What potential approaches exist in current literature to address the issue of limited data for skin disease analysis?

RQ2 Can GANs be effectively utilised to generate synthetic images from a limited dataset, enhancing the dataset's volume and diversity for improved skin disease

analysis?

RQ3 How can synthetic Rosacea images be validated qualitatively by expert dermatologists?

RQ4 Is it feasible to generate synthetic data from a limited number of samples using 3D modeling, with precise control over the granular deformations caused by Rhinophyma?

RQ5 Can the classification models trained using synthetic images, derived from a 3D environment perform well when tested on real-world data?

1.6 Contributions

In this thesis, we tackle the issue of limited data availability for machine learning and deep learning applications, particularly in the computer vision domain. This limitation has considerable implications for both academic research and practical applications. Our proposed solution involves utilizing synthetic visual data to supplement available resources, thus bolstering the efficacy of deep learning algorithms. Specifically, we delve into two distinct image synthesis methodologies, namely Deep Generative Modelling and 3D Parametric Modelling, applied to the clinical real-world images of Rosacea skin condition.

Given the scarce nature of Rosacea datasets, this study's primary objective is to explore and contribute to strategies for working with limited data in the context of state-of-the-art deep learning and computer vision algorithms. The detailed contributions are enumerated below:

C1 A comprehensive and critical examination of the literature has been undertaken, wherein we investigate strategies to manage limited data. The primary goal of this review is to uncover potential methodologies that could mitigate the issues surrounding data scarcity and enhance diagnostic capabilities using small datasets. Given the challenge of data scarcity, various techniques such

as Generative Adversarial Networks, Meta-Learning, Few-Shot Classification, Federated Learning and 3D Face Modelling are explored and discussed. Additionally, we delve into the existing research based on the skin conditions studied, the volume of data used, and the choices made in implementation. This comprehensive review acted as a fundamental resource and provided significant inspiration for the subsequent research contributions made in this thesis.

C2 In what we believe to be a novel contribution, a limited dataset comprising 300 full-face images of Rosacea is leveraged for synthetic image generation. The study demonstrates the impact of fine-tuning a state-of-the-art model and varying experimental settings on the fidelity of Rosacea features. Extensive experimentations illustrated that the choice of Regularization strength plays a crucial role in achieving high-fidelity representations in limited data conditions. This approach facilitated the generation of 300 high-fidelity synthetic full-face images, marked by Rosacea features, which can potentially be employed to enhance the available Rosacea face dataset. We have made the Synthetic Rosacea dataset, named “synth-rff-300”, publicly available on GitHub link provided in chapter 3.

C3 Part of “synth-rff-300” was utilized for qualitative evaluations by three expert dermatologists and twenty-three non-specialist participants. These qualitative evaluations indicate how realistic the characteristics of Rosacea in the generated images are. The results of the qualitative evaluations sometimes contradict the results of the quantitative evaluations using metrics such as KID and FID. We critically analyse the quantitative evaluations and the validation metrics(s) used based on 10 conducted experiments and we emphasize that relying solely on these quantitative validation metrics may not be sufficient for evaluations in the computer-aided medical image diagnosis field.

C4 For the first time, we have successfully generated synthetic data by leveraging

a limited dataset of 268 clinical images, each representing various stages of Rhinophyma. To achieve this, we utilized a parametric 3D face model which played a crucial role in the synthetic data generation process. The 3D models are carefully designed to represent various deformations through a large range of stages representing realistic patterns of Rhinophyma which may exist but have not been photographed in the real-world. Utilizing rendering techniques, we successfully generated 2000 distinct deformations of a Rhinophyma-affected nose. Subsequently, images of each of these Rhinophyma noses were captured from ten perspectives, leading to a comprehensive collection of 20,000 images. To encourage further research in this field, we have made both the synthetic dataset and the associated 3D models publicly accessible on Zenodo and GitHub links provided in chapter 4.

C5 The synthetic Rhinophyma dataset has been employed to train deep learning-based classification models. These models were subsequently tested on a real dataset consisting of 220 curated and rigorously preprocessed Rhinophyma images, with a particular focus on accentuating important Rhinophyma nose features for model training. Following this, GradCAM [64], was utilized, enabling us to determine the specific image regions the model concentrated on during the decision-making process. This approach effectively showcased the crucial role of the synthetic dataset during the validation process.

In forthcoming research endeavours, the strategies and methodologies discussed herein may serve as foundational frameworks, particularly when confronting the challenge of limited visual data availability for specific disease cases. This can facilitate the advancement of investigation and development processes in disease diagnostics, paving the way for more efficient and effective solutions.

1.7 Thesis Structure

The following chapters of this thesis are structured as follows.

1. Chapter 2 offers a comprehensive review of the existing literature on computer-aided skin disease analysis. This analysis predominantly leverages image datasets and employs sophisticated machine learning and computer vision techniques. A substantial portion of the content within this chapter has been previously published in IEEE Access.
2. Chapter 3 elucidates the process of synthetic data generation for Rosacea subtype 1 and 2, accomplished using GANs models. This chapter comprises an in-depth account of the data collection, data preparation, conducted experiments, their corresponding quantitative results, analysis, and high fidelity data generation. In the second part of this chapter, a qualitative analysis by human evaluators is conducted on synthetic Rosacea images, followed by a critical discussion, and future directions.
3. Chapter 4 presents the procedure for synthetic data generation for Rosacea subtype 3, Rhinophyma. It covers the data collection, design choices of 3D models, rendering set up of 3D models, synthetic data generation, dissemination, and value. Furthermore, this chapter presents the application of deep learning models for Rhinophyma classification. It describes the preparation of synthetic data obtained from the 3D models, along with the methodology for both synthetic and real data preparation. This demonstrates the practical use of synthetic data in real-world scenarios, using the limited set of real-world Rhinophyma data available for the study.
4. Chapter 5 discusses the implications of the research findings by concisely summarizing the salient points and suggesting possible avenues for future research, aligned with the thematic focus of the conducted research.

Chapter 2

Background and Related Work

Skin is the largest organ of the human body which plays an important role in protecting the body from harsh chemical and environmental conditions. Skin diseases affect one third of the world's population [63]. According to a report published by the National Centre for Biotechnology Information (NCBI) in 2017, skin diseases are the fourth leading cause of non-fatal diseases worldwide [64]. Skin diseases cause discomfort in day-to-day life. They get worse with time, reduce productivity in the daily regime and, if not treated at the early stage, can be deadly. Skin diseases are not only a problem for individuals but for the world population posing an increasing economic threat to national healthcare systems worldwide [65]. According to one of the latest survey report published in 2013 by European Dermatology Health Care, the 10 countries with longest waiting times for regular dermatological visits from 40 days to 133 days are: Germany, Malta, Austria, Luxembourg, Sweden, Poland, Norway, UK, Slovenia and Ireland [66]. Likewise, there are only a few dermatologists per 100,000 population in many countries. Table 2.1 highlights the limited number of dermatologists in six different countries gathered from the official sources. Given the low number of dermatologists and long waiting times, it is essential to expand the scope of skin treatment through computer-aided diagnosis.

To complement the work of qualified dermatologists, skin disease diagnosis using Computer Vision and Machine Learning is important in contributing to the early diagnosis process performed by healthcare professionals such as General Practition-

Table 2.1: Number of dermatologists in six different countries. This table quantifies and compares the number of dermatologists available in six different countries, expressed per 100,000 population.

Country	No. of Dermatologists per 100,000 population	Source
United Kingdom	1.4	British Association of Dermatology, 2013 [67]
Ireland	1	Health Service Executive (HSE), Ireland, 2014 [68]
Canada	0.47 (rural) 1.96 (urban)	Royal College of Physicians and Surgeons of Canada, 2019 [69]
USA	3.4	Journal of American Medical Association (JAMA), American Academy of Dermatology (AAD), 2016 [70]
Australia	1.9	Australian Government, Department of Health, 2016 [71]
China	1	Chinese Medical Journal, 2019 [72]

ers and Dermatologists. From the early 90s, dermatologists have been collectively working via digital platforms to communicate and diagnose skin diseases of patients by utilising skin disease images and additional health data. In medical literature, this technique of collecting, monitoring, storing, and sharing data in order to help diagnose skin conditions is termed “Teledermatology” [73].

There are various ways of diagnosing skin diseases through imaging. The three most common kinds of skin image data are: histopathological, dermoscopic, and clinical images. A few existing studies on skin disease diagnosis using traditional machine learning algorithms have been done using histopathological images for cancerous skin conditions. Most of the work on skin diseases has been done using dermoscopic images, primarily cancerous skin lesions. However, only a few studies have been done on clinical images of common and chronic skin conditions such as acne, rosacea, eczema, lupus, seborrheic dermatitis, and a few other conditions. Hence, there is a need for attention to these diseases in medical image analysis using advanced machine learning and computer vision techniques. However, there are specific challenges to be dealt with due to the nature of these diseases and the

availability of datasets. For example, a specific skin condition called rosacea will be looked at in this review. In the case of rosacea (as for the other conditions), there is only a limited amount of image data available. For Machine Learning and Computer Vision techniques, especially when applied to skin diseases and conditions like rosacea, having a substantial amount of diverse data is crucial for training models that can accurately detect and diagnose these conditions. However, obtaining comprehensive datasets for such specific medical conditions can be challenging. In this review, we delve into various existing approaches that have utilized datasets from different skin diseases. Specifically, the following approaches are examined to address the challenges posed by limited data availability in the context of skin diseases.

1. Data Augmentation, i.e. generating synthetic data with slight modifications to complement the real data.
2. Transfer learning and fine-tuning i.e. adapting a neural network model, which has been pre-trained on another much larger dataset, to classify rosacea.
3. Generative Adversarial Networks (GANs) i.e. generating high quality synthetic faces with rosacea.
4. Meta-Learning and Few-Shot classification i.e. learning faster with fewer examples.
5. Federated Learning i.e. employing a collaborative model training approach where multiple entities contribute to learning a shared model while keeping their data localized, enhancing privacy and data utilization.
6. 3D Morphable Face Models i.e. creating a 3D model of human face with various subtypes of rosacea from a set of 2D images.

In this literature review, we are investigating skin diseases, especially the importance of rosacea diagnosis using machine learning and computer vision. In Section 2.1., skin diseases, types of medical diagnosis, types of images used in computer

vision and machine learning tasks for diagnosis of skin conditions are discussed. In Section 2.1.1., the motivation for rosacea image analysis is discussed. Section 2.2 is focused on a few existing studies carried out on rosacea using machine learning and computer vision. In Section 2.3, we discuss how the amalgamation of big data, deep learning and computer vision has brought some breakthroughs in the field of medical diagnosis. This discussion is followed by the challenges of having a smaller dataset in the field of medical diagnosis and how to leverage a smaller dataset using various techniques of machine learning and computer vision. Hence, in section 2.4, we discussed various publicly available datasets.

Section 2.5 provides a brief overview of skin disease analysis from traditional machine learning and computer vision techniques to the modern deep learning algorithms. Section 2.5 contains four subsections in which four techniques in machine learning and computer vision are discussed, a few existing studies using Data Augmentation and Transfer Learning in Section 2.5.1, Generative Adversarial Networks in Section 2.5.2, Meta-Learning and Few-Shot Classification in Section 2.5.3, Federated Learning in 2.5.4, and 3D Face Modelling in Section 2.5.5. The challenges and major takeaways are mentioned in each sub section of these four techniques. Furthermore, based on a few key points from the literature review, the implementation possibilities of GANs, Meta-learning, and 3D Face Modelling in the limited data scenario are discussed in Section 2.6. Based on the options for implementation, some future directions are recommended in Section 2.7.

2.1 Digital Imaging in Dermatology

Skin diseases are one of the most challenging fields in medical diagnosis due to their observational and analytical complexities. Diagnosis of skin diseases requires years of experience and expertise. Skin diseases are diagnosed visually, with an initial screening followed by dermoscopic analysis, biopsy and histopathological analysis. However, this process of diagnosis is time-consuming and costly. Chronic inflammatory skin diseases which may not be fatal in most situations, may still need lifelong

engagement with dermatologists. Chronic skin conditions need regular check-ups, up to date medications, surgical or laser treatments, if required. This way of treatment requires a significant amount of time and are often costly. However, detecting chronic skin conditions at the early stages allows for an intervention and prevention of further complications. Several technically advanced hospitals in the world follow a dynamic process for treating skin diseases. This requires patient record-keeping, including images of the skin diseases and basic information about the patients, which helps monitor the progress of the treatment over time.

One of the traditional and common techniques to collect patient data for diagnosing skin diseases is dermoscopic imaging. Dermoscopic images are collected through high quality magnifying lenses (mainly through a dermatoscope) with powerful lighting. The most common types of images captured by the dermatoscopes are micro and macro images of individual lesions which lack the anatomical details of the body. In the context of dermoscopic imaging, 'micro' images refer to highly magnified photographs of the skin, capturing detailed views of individual skin lesions at a close range. These images provide a close-up view of the skin's surface, revealing textures, colors, and patterns that are not visible to the naked eye. Micro images are particularly useful for examining the fine details of skin lesions, aiding in the identification of specific features indicative of various conditions like malignant or benign lesions. On the other hand, 'macro' images capture a broader view of the skin area, including multiple lesions or larger segments of skin with less magnification compared to micro images. While they provide less detailed information on a per-lesion basis, macro images are useful for understanding the broader context of the skin's condition, including the distribution and general appearance of lesions over a larger area.

Further, these dermoscopic images are captured and examined by specialist dermatologists. Hence, dermoscopic images are very useful when diagnosing individual skin lesions such as malignant and benign lesions on the body. However, it is not possible to capture dermoscopic images for every skin condition at the initial stage

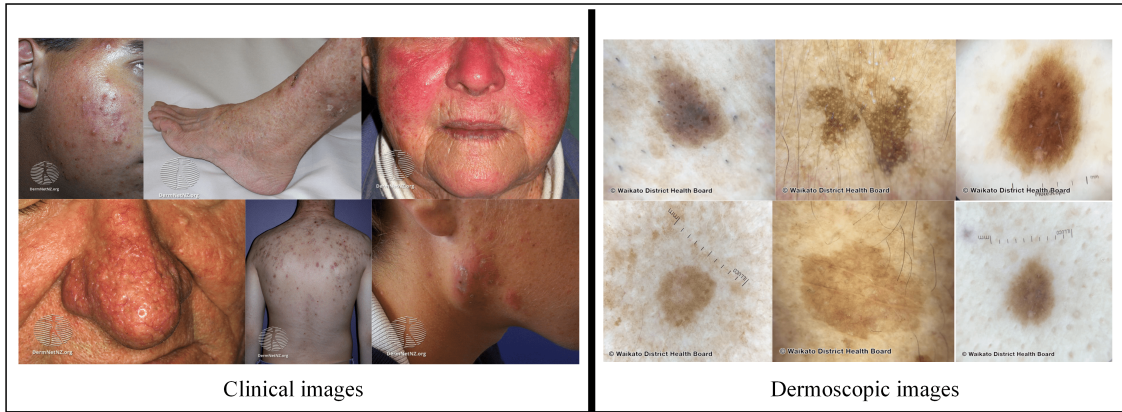


Figure 2.1: Clinical images vs. Dermoscopic images. The images used in this figure are taken from DermnetNZ [75].

of the disease. For this purpose, using a good quality digital camera or smartphone can facilitate capturing images of common skin diseases making smart phones or other digital photographic devices an accessible alternative for capturing skin conditions at the early stage of the diagnosis. The skin images captured by smartphone or other photographic devices are referred to as “clinical” images in the world of medical science research. As a result, clinical images are gaining popularity in skin disease diagnosis [74] and there are many medical research platforms that encourage collecting clinical images. Fig. 2.1, illustrates a few samples of clinical and dermoscopic images.

2.1.1 Motivation for Rosacea image analysis

Rosacea is a chronic facial skin condition that goes through a cycle of fading and relapse [14]. The frequency of fading (remission) and relapse (exacerbation) varies widely among individuals and is influenced by several factors including environmental triggers, stress, lifestyle, and individual skin type. Generally, patients may experience flare-ups lasting weeks to months followed by periods of remission where symptoms significantly lessen or disappear. On average, individuals with rosacea may experience varying degrees of flare-ups several times a year, with some reporting more frequent or prolonged episodes. The specific pattern and frequency are highly individualized, making regular monitoring and personalized treatment essen-

tial. It is also a cutaneous vascular disorder [15]. It is a common skin condition in native people from northern countries with fair skin or Celtic origins [16]. Rosacea is often characterized by signs of facial flushing and redness, inflammatory papules and pustules, telangiectasias, and facial edema. Rosacea's symptom severity varies greatly among individuals [18]. In the medical diagnostic approach, rosacea is classified into four subtypes – Subtype 1 (Erythematotelangiectatic rosacea), Subtype 2 (Papulopustular rosacea), Subtype 3 (Phymatous rosacea) and Subtype 4 (Ocular rosacea). Each subtype is diagnosed based on the severity of condition e.g. mild, moderate, or severe [14][20]. This variability in the cycle of rosacea symptoms and subtypes underscores the importance of early diagnosis and ongoing monitoring through computer-aided methods. Advanced imaging and analysis techniques can help identify and predict flare-up patterns, enabling timely interventions.

Besides the clinical complications, rosacea can affect patients' overall wellbeing, social life and work life. According to a survey carried out by National Rosacea Society of Canada, among 700 patients with rosacea in the working group, 66% were affected in their professional interactions, 33% had cancelled or postponed business meetings, 28% had missed work, 28% felt rosacea may have negatively influenced their chances of a promotion. In another survey with 660 patients with severe cases of rosacea, 86% of participants reported that they had to limit their social lives due to rosacea [76].

'British Association of Dermatologists' reported that rosacea is a facial dermatosis and therefore easily visible. It can cause extreme discomfort to those who suffer from it[20]. According to another study conducted by Spoenclin *et al.*[77] based on data collected in the period of 1995-2009, rosacea was diagnosed in 80% of cases after the age of 30 years, in which 61.5% patients were women.

According to the 'Acne and Rosacea Society of Canada' more than 3 million Canadians suffer from rosacea[78]. It is anticipated to become one of the most common health problems in Canada. One study in Sweden found that women with rosacea are more likely to experience migraine headaches than those with healthy

skin conditions[79]. In 2019, according to a Market Analysis Report published by Grand View Research[80], the rosacea treatment market will be worth \$2.6 billion by 2025, proving one of the fastest growing drug classes. In April 2021, a report published by the National Rosacea Society[23], states, “New treatments continue to expand therapy options, but a cure remains elusive”. The reasons behind the continued growth of rosacea cases in recent years are: lack of awareness in dermatologists, misdiagnosis, cost of treatment etc.[23].

In a recent article published by US News Health, “rosacea is often misdiagnosed, and many don’t seek treatment because they don’t realize it’s rosacea,” says Jeffrey Fromowitz, MD, FAAD, a dermatologist based in Boca Raton, Florida[24]. Frequent news on rosacea awareness and treatment appears regularly in the Irish Times[81], Irish Examiner, and a few other newspaper organisations in Ireland. This indicates the global scale of the problem of rosacea. As the concern rises, the treatment of rosacea is not only the responsibility of expert dermatologists, but Machine Learning can also be a potential pathway towards the early diagnosis of rosacea with state-of-the-art methodologies. A fast, accurate and low-cost assistive diagnostic system could significantly contribute to medical treatment plans, particularly in developing countries. Early and accurate detection of skin lesions, inflammation and facial skin conditions, such as rosacea, is vital for developing precise and effective treatment and medication. In this review we provide a critical literature review and an analysis on skin disease diagnosis using various methodologies of machine learning and computer vision.

2.2 Related work on Rosacea

In recent years, medical image diagnosis has progressed rapidly due to the advancement of Artificial Intelligence (AI) models and the availability of a large amount of data provided by medical professionals. Thanks to the advanced machine learning and deep learning techniques in computer vision, different types of disease diagnosis have become very widespread in the scientific and medical research community. An

extensive amount of work has been done on skin cancer diagnosis. According to a study published by Stanford University in 2017[82], Dermatological (dermoscopic) images play an important role in diagnosing skin cancer using Deep Convolutional Neural Networks (DCNNs)[83]. The work by Esteva *et al.*[82] suggests that the diagnosis technique could be used outside the clinic as an initial screening step for cancer to a level of competence comparable to 21 board-certified dermatologists. Since then, computer vision and deep learning research has attracted a lot of attention for skin cancer lesion classifications by proposing various kinds of state-of-the-art methodologies and techniques. However, most of the work on skin disease analysis and classification so far has been done is on dermoscopic images, in which a particular region of interest of the skin is focused on, as shown in Fig. 2.1; while there are very few studies on facial skin conditions such as rosacea, rosacea acne, eczema, psoriasis lupin and other related skin conditions. Table 2.2 presents an overview of studies conducted on rosacea along with other skin conditions.

As it can be seen from Table 2.2, most of the studies which have been carried out on rosacea and related skin conditions using machine learning and computer vision/deep learning algorithms date from the year 2019 and onwards. Most of the works which have shown great results using deep learning have used at least nearly 10,000 images. A few studies conducted by Thomsen *et al.*[84], Zhao *et al.*[85], Wu *et al.*[86] and Zhu *et al.*[87], employ a significant quantity of data. However, the datasets used in these studies are entirely confidential. Hence these studies are not entirely reproducible and therefore there is a motivation for other researchers to try to deal with the skin disease problem using limited data. Most of the work done so far using transfer learning and data augmentation has used weights pre-trained on ImageNet[3], which is considered a non-medical dataset. However, these studies provide a useful insight into a few common techniques which can be applied in this research. A few studies have shared their GitHub repository, which may provide references for publicly available datasets.

Goceri [88] presented a novel modified Mobile-Net architecture [89] along with

a mobile app with user-friendly interface. In this work, 725 images of seborrheic dermatitis, rosacea, hemangioma, psoriasis and acne vulgaris were used for classification tasks. There were 145 images in each disease class. The modified Mobile-Net model was developed based on the original Mobile-Net architecture but with the receptive field expanded, with dilated convolution and combined hybrid loss functions. The experimental results in this study have shown that the proposed modified Mobile-Net[88] has outperformed other network architectures for each disease class.

Thomsen *et al.*[84] presented a dataset and a classification task with 5 categories of skin conditions i.e. Psoriasis, Eczema, cutaneous t-cell lymphoma, acne and rosacea. As part of the pre-processing before the classification task, K-means clustering was used to remove the noise and unnecessary details from the images. Four types of modified VGG-16 CNN architectures [90], either incorporating or not incorporating the Spatial Transformer Network (STN), are depicted in Table 2.2. According to the results discussed in this study, these 4 types of VGG-16 architecture performed differently in terms of Area Under the Curve (AUC) and accuracy scores for each disease class. However, VGG-16P is proven to be the best performing model after the performance measured through specificity, sensitivity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Additionally, the overfitting due to the small datasets and selection bias for acne and rosacea is an issue and was discussed in this study.

Goceri [91] proposed a segmentation method called Fully Automated Detection of Facial Disorders (F-ADFD). This method has shown better segmentation accuracy, specificity, and precision due to active contouring which is set automatically using a binary image that is obtained with a K-means clustering after denoising and intensity normalization steps. Among 10 Deep Neural Net (DNN) architectures that were used in this work, DenseNet201[92] with modified loss function (cross-entropy and Tversky similarity) was claimed to have shown results with maximum accuracy (95.24%) and minimum loss (0.5). The second highest performance was obtained by InceptionResNet-v2 [93]. This study suggests that DNN techniques can

extract features automatically at low, middle and, high levels by increasing depth and by performing classification for skin lesions. All the DNN architectures used in this work were pretrained on ImageNet [3].

Zhao *et al.*[85] carried out a study on three subtypes of rosacea lesions i.e. Erythematotelangiectatic rosacea (ETR), papulopustular rosacea (PPR), and phymatous rosacea (PhR). The study reported classification accuracies of 83.9%, 74.3%, and 80.0% for ETR, PPR, and PhR, respectively, when distinguishing each subtype from the others. Beyond rosacea, the study encompassed additional skin conditions resembling rosacea, including acne, facial eczema, seborrheic dermatitis, lupus erythematosus, chronic solar dermatitis, corticosteroid-dependent dermatitis, and lupus miliaris disseminatus faciei. A total of 24,736 images were utilized for the analysis, though the study did not disclose the source of these images, nor is the dataset publicly accessible. Furthermore, the authors highlighted the necessity of investigating deep CNN decision-making processes. Understanding these processes is crucial for enhancing the models' accuracy and specificity in disease detection, ultimately improving diagnostic capabilities in dermatological conditions.

Wu *et al.*[86] performed a classification among psoriasis (Pso), eczema (Ecz), Atopic dermatitis (AD) and Healthy skin. This work did not use any rosacea images, and instead relies on hand and facial images. The study involved 4,740 images collected from the Department of Dermatology, The Second Xiangya Hospital, Central South University, China. However, this dataset is confidential. Google's EfficientNet-b4 [94] was used with an extra 7 auxiliary classifiers at the end of each intermediate layer to make the model learn classification information from different levels of features. This work was built as a smart phone mobile application.

Zhu *et al.*[87] performed a classification among 14 skin diseases with 13,603 images labeled by two dermatologists with a minimum of 5 years of experience. This data was collected from the Department of Dermatology, Peking Union Medical College Hospital, China, from April 2016 to April 2020. In this work, EfficientNet-b4 [94] was used with pre-trained weights from ImageNet. There were 14 classifiers with

14 output neurons used instead of the final fully connected classification layer of the network. This modified model of EfficientNet-b4 is compared with Inception-v3 [95], ResNet-101 [96] and the original EfficientNet-b4 [94]. The comparative outcomes are measured by AUC, ROC, Sensitivity, Specificity and Accuracy. The modified EfficientNet-b4 has outperformed other CNN models with an AUC of 0.985 and with the highest ROC. The performance of modified(proposed) EfficientNet-b4 was also compared with dermatologists, in which performance is measured using the Kappa coefficient [97]. This performance measure comparison showed that the diagnosis of Rosacea by dermatologists is significantly better than the proposed model. In comparison, the diagnosis of viral warts by the proposed model was significantly better than the dermatologists.

Aggarwal [98] looked at 5 skin conditions: acne, atopic dermatitis, impetigo, psoriasis, and rosacea. A total of 938 images were considered for classification using Inception-v3 with pre-trained weights of ImageNet [3]. Data Augmentation was incorporated during the training process to reduce the possibility of overfitting. The performance of the model was measured through sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Mathew's correlation coefficient (MCC), and F1 score. A comparative result illustrated how each model performed with and without data augmentation. From the confusion matrix, the performance scores for rosacea are relatively low (0.60 with data augmentation) compared to the other four skin conditions. The number of images of rosacea considered in this study was 90.

Binol *et al.*[99] presented a study using rosacea images collected from the Division of Dermatology at Ohio State University. There were 41 images collected using a DSLR camera, in which images were taken from left, right, front and upsides of the faces. There were two CNN models considered for classification i.e. Inception-ResNet-v2 [93] and ResNet-101 [96] with the pre-trained weights from ImageNet [3]. A few pre-processing tasks were performed, such as creating labelled patches on the facial images. These patches were labelled by expert dermatologists based on the

anatomical details of the face. The anatomical parts of the face regions more than 75% affected by rosacea were labelled as positive rosacea patches and less than 25% rosacea affected regions were labelled as negative rosacea patches. The patches with various resolutions such as 64×64 , 128×128 , 256×256 were obtained for the data augmentation and transfer learning process. Hence there were nearly 65,649 tiles with different resolutions crafted on the 41 images. The accuracy of the models was measured using Dice Coefficient and false-positive rate. A specific kind of post-processing was proposed in this work called Anthropometric Post-Processing (APP) with a landmarks-based Region of Interest (ROI) mask. The Inception-ResNet-v2 [93] with APP provided a higher performance score compared to other models such as ResNet101 [96] and Bag of Features with Support Vector Machine (SVM).

Xie *et al.*[100] presented a dataset and classification task for 541 skin conditions. The image dataset was collected using 4 types of digital cameras, and these images were annotated by 20 professional dermatologists from the Xiangya Hospital of Central South University, China. However, in this study, 80 categories of skin conditions were considered for classification. The disease categories considered had more than 100 images and the categories with more than 1000 images were discarded to keep a balance during the classification process. There were 4 types of CNN architectures considered for the task i.e. InceptionResNet-v2 [93], Inception-v3 [95], Densenet121 [92] and Xception [101]. According to the results drawn from the confusion matrix, InceptionResNet-v2 outperformed the other three CNN architectures with 0.764 accuracy.

2.3 Learning - from “Big Data” to “Small Data”

The concept of visual data – ‘image datasets’ started gaining popularity in 1999 through the release of an official standard database i.e. the MNIST database (Modified National Institute of Standards and Technology database) by Yann Le Cun and his colleagues[102][103]. The MNIST database is a collection of handwritten digits. It has a training set of 60,000 example images and a test set of 10,000 images.

In 2009, Deng *et al.*[3] introduced ImageNet, one of the largest image datasets available containing around 3.2 million images. Based on the numbers recorded on the ImageNet homepage, there are more than 14 million images in the dataset with just over 21 thousand synsets (groups/classes).

The real-world artefacts which humans can recognise have now become recognisable by computers through efficient algorithms and large sets of images, which was a difficult task a decade earlier. These advancements have become possible due to the availability of large volume datasets like ImageNet [3], which can be fundamentally called ‘Big Data’. Having bigger datasets is one of the key prerequisites for Deep Learning models to perform well.

2.3.1 Big Data in computer-aided medical diagnosis

Medical image analysis using deep learning has become popular among research communities due to the collective concept of ‘Big Data’. According to the studies reviewed in this work, the volume of the data used for medical diagnosis is three to four times smaller than the number of images in ImageNet [3].

One of the influential works by Esteva *et al.*[82] on skin disease analysis with dermoscopic images for diagnosing skin cancer created a new trend for skin disease analysis using deep learning and computer vision. In total 129,450 clinical images were used in [82] to train a deep convolutional neural network to classify the most common deadliest skin cancer. Advances in medical image analysis techniques have also been used for other conditions. For instance, Ting *et al.*[104] utilised a dataset of 494,661 retinal images to diagnose Diabetic Retinopathy and related eye diseases from visual scans. In order to detect bone fractures in radiographs, a deep learning model trained on 135,845 radiographs of a variety of body parts was proposed with a diagnostic accuracy similar to that of senior subspecialized orthopedic surgeons [105]. This work claims that, given enough training data and a suitably designed model, it is possible to detect any condition on radiographs that a human clinician could identify. A deep learning model, DeepSeeNet [106] was developed to classify

patients with Age-related Macular Degeneration (AMD). DeepSeeNet was trained on 58,402 training images and 900 testing images collected from 4549 participants. In another attempt, a deep learning model was trained for automatic Magnetic Resonance Imaging (MRI) cardiac multi-structure segmentation and diagnosis[107]. It used the “Automatic Cardiac Diagnosis Challenge” dataset (ACDC), the largest publicly available and fully annotated dataset for Cardiac MRI (CMRI) assessment. The dataset contains CMRI recordings obtained from 150 devices, with reference measurements and classification from two medical experts.

2.3.2 Importance of the ‘small datasets’ in medical diagnosis

Although deep learning models have exhibited prodigious performance in computer vision tasks such as automated diagnosis of medical conditions (diabetes, retinopathy, bone fractures, age-related macular degeneration, cardiac MRI and skin cancer, etc.), they heavily rely on a large volume of the labelled dataset [108]. While these models are helping to achieve breakthrough state-of-the-art performance, their accuracy downgrades severely on datasets with only a few labelled instances [109]. In various cases of rare diseases, it is difficult to acquire and annotate an adequate number of samples for large-scale training assignments. As a consequence, these models end up with poor generalization for novel classes given the low number of instances per class. Training a deep learning model in a low-data scenario results in a long-tailed and imbalanced classification, which is a challenging task in both computer vision and medical imaging [8] [9] [10].

Acquiring a large amount of labelled data for real-world problems including medical imaging and clinical diagnosis, is an exhaustive and expensive task. Therefore, it would be highly desirable to improve the learning strategies with the limited amount of available data. In this work, we explore the strategies to leverage the existing approaches for learning from limited data and build generalised models from relatively small samples. The main motivation to deal with the limited data in this research is the limited availability of datasets for rosacea.

Table 2.2: Comprehensive Summary of Research Studies on Rosacea and Associated Skin Conditions. This table encapsulates a detailed survey of research efforts undertaken to understand and classify Rosacea alongside various other skin conditions. It outlines each study by the author(s), publication year, the specific skin diseases addressed, dataset details including source and availability, volume of data used, the approaches taken to address the problem, methodologies applied, performance metrics attained, and deployment strategies.

Author/ Index/ Year	Skin Disease	Dataset/ Source/ Availability	Dataset Volume	Problem Approach	Methodology	Performance Measures	Deployment
Evgin Goceri [88], 2021	seborrheic dermatitis, rosacea, hemangioma, psoriasis and acne vul- garis.	DermWeb, DermNet Der- matoweb, DermQuest	Total no. of Images = 725, No. of images for each class = 145	Segmentation, Classifica- tion prob- lem, Mobile App	Pretrained on Ima- geNet weights: SqueezeNet, Shuf- fleNet, MobileNet, RMNv2, MobileNet- V2, Light Weight Efficient Network (LWEN), Look- Behind Fully-CNN (LB-FCN), Light CustomNet2, Modi- fied MobileNet-V2	Accuracy, Specificity, Sensitivity, Precision, F1 Score, Matthew's correlation coefficient (MCC)	32 GB RAM, Intel i9-9900 pro- cessor unit (3.10 GHz) and 64-bit Windows- 10. Python 3.6, Java, Android Studio (ver- sion 3.6.1), TensorFlow
Thomsen <i>et</i> <i>al.</i> [84], 2020	acne (581) rosacea (1606), psoriasis (6,545), eczema (5,350) and cuta- neous t-cell (2,461)	Department of Dermatology, Aarhus Univer- sity Hospital (AUH), Den- mark (Confid- ential Dataset)	Total im- ages = 16,543 Total number of patients included in the study = 2,342	Classification problem, Region of interest using STN	Pre-trained on Ima- geNet weights with VGG-16 (VGG- 16P), No pretrained VGG-16 (VGG-16N), Spatial Transforma- tion Network (STN) with a Pre-trained VGG-16 Model (VGG-16PS), No pretrained VGG-16 with STN (VGG-16 NS)	AUC, sen- sitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV), Accuracy	N/A
Evgin Go- ceri[91], 2021	seborrheic dermatitis, rosacea, he- mangioma, psoriasis and acne vulgaris.	DermWeb, DermNet Der- matoweb, DermQuest.	N/A	Segmentation, Denoising, Intensity normaliza- tion, Fully Automated Detection of Facial Disorders (F-ADFD) method, Classifica- tion	Pretrained on ImageNet, VG- GNet16, VGGNet19, Google-Net, Incep- tionV3, Xception, ResNet18, ResNet50, ResNet101, In- ceptionResNetV2, DenseNet201with modified loss func- tion (Cross-entropy and Tversky (Tv) similarity)	For segmentation- Area Er- ror Rate (AER), For classification- Accuracy, Precision, Specificity, F1 score, MCC	Intel Core i7, 8GB DDR4 RAM, 3.6 GHz CPU, All networks have been trained using MATLAB (R2019b) on the same computer

Continuation of Table 2.2

Author/ Index/ Year	Skin Disease	Dataset/ Source/ Availability	Dataset Volume	Problem Approach	Methodology	Performance Measures	Deployment
Zhao <i>et al.</i> [85], 2021	3 rosacea subtypes (erythematotelangiectatic rosacea, papulopustular rosacea, and phymatous rosacea) acne, seborrheic dermatitis, and eczema	NA, Confidential Dataset, Data collection devices: iPhone X, Huawei P20 and digital camera Canon Rebel 550 from 3 different angles.	Total = 24,736, rosacea = 18,647; Acne, Seborrheic Dermatitis, eczema = 6089	Feature extraction, Classification	Pretrained on ImageNet, ResNet-50, mini-batch gradient descent with a momentum = 0.9, batch size = 32, Training epochs=100, Initial learning rate = 0.0001. (If validation loss did not decrease in continuous 10 epochs, the learning rate was divided by 5), minimum learning rate = 0.000001.	Accuracy, precision, Area Under the Receiver Operating Characteristic Curve (AUROC)	N/A
Wu <i>et al.</i> [86], 2020	psoriasis (Pso), eczema (Ecz), atopic dermatitis (AD), healthy skin	Department of Dermatology, The Second Xiangya Hospital, Central South University, China, Confidential dataset	Total = 4,740 clinical images	Classification, Mobile App.	Five-fold cross-validation to validate the effectiveness, pre-trained weights on ImageNet, EfficientNet-b4 (380×380) The final fully connected classification layer was replaced with 3 output neurons. Also, added 7 auxiliary classifiers at the end of each intermediate layer to make the model learn classification information from different levels of features.	Positive rate (TPR), false positive rate (FPR), ROC, AUC, T-SNE analysis, Confusion matrix.	Pytorch 1.1. CPU - 18 Core Intel Xeon E5-2697, GPUs - 4 RTX 2080Ti NVIDIA.

Continuation of Table 2.2

Author/ Index/ Year	Skin Disease	Dataset/ Source/ Availability	Dataset Volume	Problem Approach	Methodology	Performance Measures	Deployment
Zhu <i>et al.</i> [87], 2021	14 diseases- lichen planus (LP), rosacea (Rosa), viral warts (VW), acne vulgaris (AV), keloid and hyper- trophic scar (KAHS), eczema and dermatitis (EAD), dermatofi- broma (DF), se- borrheic dermati- tis (SD), seborrheic kerato- sis (SK), melanocytic nevus (MN), he- mangioma (Hem), psoriasis (Pso), port wine stain (PWS), basal cell carcinoma (BCC).	Department of Dermatol- ogy, Peking Union Medical College Hos- pital, China, Collected from October 2016 to April 2020, Confidential dataset The annotation process was performed by 2 dermatolo- gists with more than 5-years' experience. Data collection device- Mole- Max HD 1.0 dermoscope, Digital Image Systems, Vi- enna, Austria.	Total = 13,603 derma- tologist - labeled dermo- scopic Images, Rosacea = 597 images	Classification and Clus- tering	Pre-trained weights on ImageNet, Google's Efficient- Net-b4 (380×380) The final fully con- nected classification layer was replaced with 14 output neu- rons. Also, with added 7 auxiliary classifiers to each of the intermedi- ate layer groups. t-SNE (t-distributed Stochastic Neighbor Embedding)	Area under curve (AUC), Accuracy, Sensitivity, Specificity, ROC, com- pared this model with 280 board- certificated dermatolo- gists.	Pytorch Scikit-learn 0.22.2 and Numpy 1.16.4.

Continuation of Table 2.2

Author/ Index/ Year	Skin Disease	Dataset/ Source/ Availability	Dataset Volume	Problem Approach	Methodology	Performance Measures	Deployment
Pushkar Aggarwal[98], 2019	acne (332), atopic dermatitis (92), impetigo (138), psoriasis (280), rosacea (96).	DermNet NZ, Dermatology Atlas, Hellenic Dermatological Atlas and downloaded images from the Google search results.	Total = 938	Classification	Pretrained on ImageNet -Inception v3	Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Matthew's correlation coefficient (MCC), and F1 score.	TensorFlow.
Binol <i>et al.</i> [99], 2019	rosacea lesions	Ohio State University (OSU) Division of Dermatology (Using DSLR camera), Confidential dataset	Total = 41 facial images, The size of each image is 4608×3072	Image classification problem for rosacea and non rosacea lesions.	Pre-trained on ImageNet DCNNs: Inception-ResNet-v2, ResNet-101, Data Augmentation. Anatomically directed post-processing (APP) (Anthropometric model)	Dice coefficient, False positive rate.	MATLAB R2018b using the Deep Learning Toolbox, (HPC) with 128 GB RAM and 16 GB NVIDIA Tesla P100 PCI-E GPU
Xie <i>et al.</i> [100], 2019	80 skin diseases with each class have more than 100 images. (Includes rosacea).	Xiangya Hospital of Central South University Dataset is annotated by 20 professional dermatologists, Confidential Dataset, Data collection device: SONY DSC-HX50 (350dpi), CANON IXUS 50 (180dpi), NIKON D40 (300dpi), NIKON COOLPIX L340 (300dpi).	Total = 47,075 images were obtained using 4 types of digital cameras.	Classification.	Pretrained on ImageNet. Inception-ResNet-v2 (for 80 skin diseases classification) (Max training epochs=5000, basic learning rates = 0.001, batch size = 25, optimizer=Adam, and the loss function=categorical cross entropy). Inception V3, DenseNet121, Xception for comparative analysis.	Top-1 and Top-3 accuracies can reach 0.588 and 0.764. 4-fold cross validation.	3X NVIDIA TITAN Xp.
End of Table							

Table 2.3: Comprehensive catalog of accessible skin disease datasets. This table provides an extensive overview of publicly accessible datasets pertinent to skin disease research, including Rosacea. It enumerates each dataset, detailing its name, covered disease categories or specific conditions, imaging modality employed, total volume of data, number of classes featured, specific mention of Rosacea images if applicable, accessibility status, and originating country or region.

Index	Dataset Name	Disease Categories/ Names	Imaging modality	Volume	Classes	Rosacea images	Accessibility	Country/ Region
1	7-point criteria (aka derm7pt)[110] 2019	Melanoma and non-Melanoma skin lesions	Clinical and Dermoscopic	>2000	~20	0	Public	Canada, Italy
2	Asan and Hallym Dataset [111] 2018	12 types of Skin Cancerous lesions	Dermoscopic	17,125	12	0	Partially	South Korea
3	Dermatology ATLAS [112] 1999	All kinds of skin diseases (including rosacea)	Clinical	~11,000	~550	38	Public	Brazil
4	DanDerm [113] 1995	All kinds of skin diseases (including rosacea)	Clinical	>3,000	~100	17	Public	Denmark
5	DermIS [114]	All kinds of skin diseases (including rosacea)	Clinical	~7,000	~700	49	Public	Germany
6	Dermnet Skin Disease Atlas [115]1998	All kinds of skin diseases (including rosacea)	Miscellaneous	~23,000	N/A	0	Public	United States
7	Dermofit Image Library (aka Edinburgh Dataset)[116]	Cancerous skin lesions	Dermoscopic	1,300	10	0	Under License Agreement	Scotland, UK
8	DermNetNZ [75] 2016	All kinds of skin diseases (including rosacea)	Clinical and Dermoscopic	>25,000	>2,500	~50	Public	New Zealand
9	Dermatoweb.net [117] 2002	All kinds of skin diseases (including rosacea)	Clinical and Dermoscopic	>7,300	0	45	Public	Spain
10	HAM10000 [118] 2018	Pigmented malignant and benign skin lesions	Dermoscopic	10,015	7	0	Public	Austria
11	Hellenic Dermatological Atlas [119] 2011	Common disease categories (including rosacea)	Miscellaneous	2,663	N/A	9	Public	Greece
12	ISIC [120] [121] 2016	Melanoma, seborrheic keratosis, benign nevi	Dermoscopic	>33,000	N/A	0	Public	Miscellaneous
13	MED-NODE [122] 2015	Melanoma and benign nevi	Microscopic	170	2	0	Public	Netherlands
14	MoleMap [123] [124] 2003-2015	Malignant and benign lesions	Clinical and Dermoscopic	>32,000	N/A	0	NA	New Zealand
15	PH2 Dataset [125] 2013	common nevi, atypical nevi, and melanomas	Dermoscopic	200 (80 + 80 + 40)	3	0	Public	Portugal
16	SD-128 [126] 2016	128 disease categories (Including rosacea)	Clinical	5,619	128 (>20 samples per class)	N/A	On request only	China
17	SD-198 [127]	198 disease categories (Including rosacea)	Clinical	6,584	198 (10-20 samples per class)	N/A	On request only	China

2.4 Availability of Skin Disease Datasets and Challenges

As deep learning models require a large amount of data for training, it is essential to benchmark the available datasets for skin disease analysis. In this study, we utilise datasets containing various skin disease images. These 17 datasets are also used in further sections where different types of deep Learning architectures and models are discussed in detail. The main purpose of this section is to provide an overview of the available skin image datasets, categorised by name, data source, disease category, imaging modality, dataset volume, number of classes, data accessibility, and the frequency of rosacea in existing datasets.

The Table 2.3 presents a compilation of accessible skin disease datasets that vary widely in their scope and characteristics. Notably, the datasets encompass a range of skin conditions, from common diseases to more specialised subtypes like melanoma and rosacea. The diversity of conditions covered allows for the potential development of comprehensive diagnostic models that can differentiate between various skin diseases.

The volume of images across these datasets varies significantly, from as few as 170 in the MED-NODE dataset to over 33,000 in the ISIC 2016 collection. The number of classes represented also varies, affecting the specificity of potential classification models. It is important to note that while larger datasets provide a broad scope for training, they require careful handling to prevent overfitting and to ensure that models remain generalizable.

The predominance of cancer-related images in the accessible datasets, as highlighted by the table, reflects the intensive research focus on skin cancer due to its severity and higher incidence rates. These datasets often feature a well-established, standardised collection of images, particularly for melanoma, which facilitates the development of automated diagnostic tools. The accessibility of these images supports widespread research efforts aimed at early detection and treatment, which are critical in cancer care.

In contrast, images of rosacea are less commonly included and are not as standardised or accessible. This disparity likely stems from rosacea being a non-fatal condition and therefore possibly less prioritised in research funding and data collection initiatives. A critical observation from the table is the inconsistent representation of rosacea images across the datasets. Some datasets, such as DERMIS and

Dermatology Atlas, include a notable number of rosacea cases, which is crucial for studies like ours focusing on this condition. The variability in rosacea representation highlights the challenge in sourcing sufficient data and underscores the need for targeted data collection efforts. Consequently, the limited representation of rosacea in these datasets can impede the development of specialised diagnostic models for this condition.

For researchers and medical professionals focusing on rosacea, this imbalance necessitates seeking out specific datasets that contain a sufficient number of rosacea images or investing in the creation of new, comprehensive collections. Ensuring the inclusion of a diverse range of rosacea presentations in such datasets is crucial for developing effective diagnostic models that are robust and generalizable across the spectrum of the condition’s manifestations.

The imaging modalities used, including clinical, dermoscopic, and microscopic images, offer different insights into skin conditions. Clinical images provide a general view, while dermoscopic images allow for the examination of skin lesions in greater detail, which can be particularly useful for conditions like skin cancer images that have distinct visual patterns. The diagnosis of rosacea typically requires examination of full-face images or images where substantial portions or specific localities of the face are visible. This approach is necessary due to the nature of rosacea, which often presents with symptoms such as redness, visible blood vessels, and swelling across different areas of the face. Unlike some skin cancers that may be identified by analysing individual lesions, rosacea’s diagnosis depends on assessing the pattern and extent of these symptoms over larger facial regions.

As shown in Table 2.3, there are only about 200 images of rosacea in publicly available datasets. Among the available images of rosacea, there is only a small number of images with the full-face visibility. Compared to the studies published based on skin cancer images, there is a very limited number of annotated rosacea images and that introduces a significant challenge in dataset split (train, validation and test) for training deep learning models.

Accessibility is another key factor; while many datasets are publicly available, facilitating open research and collaboration, others are restricted or require specific agreements to access. This can limit the utility of the datasets for widespread research purposes.

Geographical representation is also a point of discussion, with datasets originating from various regions such as Canada, Italy, South Korea, and China. The geo-

graphic diversity can help in developing diagnostic models that are effective across different populations, addressing the issue of dataset bias that can arise from a narrow geographic focus.

Lastly, legal and ethical considerations are implicit in the accessibility column. Datasets like the Edinburgh Dermatology Image Library require a license agreement for access, which may include ethical considerations such as patient consent and privacy protection, particularly relevant when dealing with identifiable human images.

In summary, the datasets listed in the table offer a rich resource for the development of skin disease classification models. However, the varying volume, class representation, modality, accessibility, and geographical origin of the datasets present both opportunities and challenges. For our study on rosacea, the selection of datasets with adequate Rosacea images and the appropriate imaging modality will be crucial for developing accurate and reliable diagnostic models. Hence, this chapter is focused on examining various deep learning techniques which may be applied for skin disease diagnosis and potentially suitable for dealing with a limited dataset.

2.5 A brief Overview of Skin Disease Analysis using Machine Learning and Computer Vision methods

Considerable amounts of work on skin disease classification tasks have focused on computer-aided skin cancer diagnosis and classification support systems [128, 129, 130, 131, 132]. These non-invasive [130, 131] [133] methods, such as traditional image processing techniques, have been very popular and achieved notable results for skin disease diagnosis, particularly for skin cancer. The image processing techniques have been used to perform a broad range of image pre-processing, e.g. lesion segmentation and domain-specific feature extraction followed by classification tasks. Such diagnosis tasks use a small number of datasets [131, 132, 133]. Generally, these datasets contain less than a thousand sample images. Image classification problems that utilise small datasets, do not generalise well to new images i.e. a novel category of diseases or the classes with very small datasets.

Over the past few years, medical image classification has entered a new era thanks

to the advancements of deep convolutional neural networks[83] [134], machine learning [135, 136, 137] and deep learning techniques [138, 139, 1]. These techniques do not require any hand-crafted features, but they heavily rely on high computational power [3] [140, 141, 142]. They are trained end-to-end directly from the image labels and raw pixels, with a single convolutional neural network for dermoscopic and clinical images [82].

Another principle that has attracted a lot of attention recently in the medical image analysis domain is transfer learning. The core idea behind Transfer Learning is to deal with fewer samples and is discussed in Section 2.5.1.

Another approach is GANs. There have been a few studies on using GANs [143, 40] on medical image analysis [144] which are discussed in Section 2.5.2. While exploring GANs has shown notable results, they come with a few limitations such as (1) mode collapse: when the generator collapses to map all latent space inputs to the same data and (2) instability: refers to the training process where the adversarial nature between the generator and discriminator can lead to unstable training dynamics. This might result in oscillating or diverging loss during training, leading to poor quality of generated images or failure of the model to converge on a solution. The principal causes for these phenomena are related to vanishing gradients through the optimization procedure [145]. However, when it comes to synthetic image generation of faces, a few types of GAN architectures have become successful as further discussed in Section 2.5.2.

Given that GANs come with a few limitations and advantages, there is a scope to explore other subfields of machine learning to deal with limited datasets, such as meta-learning and few-shot learning [146, 147, 148]. Meta learning approaches differ from many standard machine learning algorithms. Meta learning systems are trained by being exposed to many tasks and are tested in their ability to learn new tasks. An example of a task might be classifying a new image within 7 possible classes, given one example of each class [147]. There has been a small number of studies based on meta learning and few-shot learning applied to medical images and skin disease diagnosis which are further discussed in Section 2.5.3.

An increasingly pertinent approach in the realm of medical imaging, particularly when confronting the challenges of data scarcity and privacy concerns, is Federated Learning (FL). This method involves a collaborative, decentralized machine learning approach where multiple clients, such as hospitals or research institutions, train algorithms collaboratively while keeping the patient data localized. This approach

is particularly beneficial for skin disease classification tasks, where data may be sparse, sensitive, and unevenly distributed across sources. Federated Learning’s unique approach offers a promising solution to harness the power of collective data while adhering to strict privacy regulations and is discussed in Section 2.5.4.

Advanced machine learning have been gaining popularity in the field of computer vision due to their advantages. Nevertheless, some of the traditional approaches in computer vision can be utilised when there is only a limited amount of data available. One of these approaches leverages the technique of 3D modeling. 3D face modelling is a computer graphics technique. By using an intuitive user interface several 3D face models can be created from one or more photographs[31]. Over the years various methodologies have been developed to reconstruct 3D faces, such as 3D Morphable Models[32], Active Shape Models [149, 150, 151], Gaussian Process Morphable Models [152], deep learning based reconstructions[153, 154, 155] and 3D modeling using GANs[156, 45]. However, there is only a small amount of literature available on medical image diagnosis and facial skin image diagnosis which will be discussed in Section 2.5.5.

2.5.1 Data Augmentation and Transfer Learning

Data augmentation is a technique to artificially create a new set of training data from the existing ones by a slight modification. This is a process of modifying and expanding the data through various geometric transformations and image processing tasks such as such as rotations, flips, and zooms, we can simulate the variability that a model would encounter in a real-world scenario. For images like those of faces with rosacea, the geometry of the faces and the positions of affected regions on the faces are different in each image. These are known as positional biases. Data augmentation can work well with the positional biases present in images, in order to increase the size and quality of training datasets, especially with a facial dataset [157].

Humans learn from their experience, which helps them understand and solve new but similar tasks quickly. A similar kind of hypothesis is applied in a handful of algorithms and techniques [158]. Transfer learning is one of the deep learning approaches in which a new task, which is in a different but related category, can be learned and improved by acquiring experience from a previously learned task. Thus, the acquired learning experience comes from different constraints such as extracting features and fine tuning the model. These constraints play an important role along

with monitoring the parameters of the model to obtain the desirable output.

Transfer learning and data augmentation are indeed distinct techniques, each with its own role in the development of computer vision models. The reason they are discussed together in this section is due to their synergistic application in many studies within the field of computer-aided diagnosis and classification. In practice, when applying transfer learning to a new problem, especially in cases where the available data is limited, it is often beneficial to augment the dataset to prevent overfitting and improve the model's ability to generalize from the learned features. Conversely, when employing data augmentation strategies, using a model that has been pre-trained on a larger, diverse dataset can provide a more robust starting point for learning the new task.

Therefore, while transfer learning and data augmentation serve different purposes, their combined use is a common and effective strategy in computer vision problems. This combined approach allows for the efficient adaptation of models to new tasks with improved performance, which is why they are frequently implemented together in the studies discussed. This integration is particularly pertinent in medical/clinical imaging, where data is often scarce and models must be highly accurate and generalizable. Hence, they are presented in this section to reflect their interconnected roles in enhancing model performance in the classification of skin diseases.

These principles are used by Esteva *et al.* [82] to demonstrate a generalizable classification of a dermatologist-labelled dataset of 129,450 images including 3,374 dermoscopy images. A GoogleNet Inception version 3 CNN architecture [95] was pretrained on approximately 1.28 million images with 1,000 categories of real world objects from the 2014 ImageNet Large Scale Visual Recognition Challenge [141]. This model was fine-tuned on a skin cancer image dataset using transfer learning [158] to achieve 93.33% accuracy.

Table 2.4 presents an overview of the state-of-the-art studies which have utilised transfer learning and data augmentation principles for skin disease analysis.

The main points that can be drawn from the Table 2.4 are:

- Most of the work done using transfer learning for skin diseases analysis starts from 2016 onwards.
- Most of the studies were conducted on subtypes of skin cancer such as malignant melanoma and benign nevi. However, only a few studies related to

rosacea were conducted so far. As seen in the Section 2.2., most of the works on rosacea and facial skin conditions are conducted from 2019 onwards.

- Most studies used a minimum data volume of 1000 images. A small number of studies have been conducted with less than 1000 images.
- The studies by Esteva *et al.*[82], Liu *et al.*[159] obtained an accuracy of 93.33% and 93% respectively with a large number of images and used InceptionNet-v3[95] and v4[93] respectively. The studies by Goceri [160], MAA [161], Cui *et al.*[162] used a small number of images to train an Inception Net[95] model. There are a few studies with a small number of datasets that obtained results by using different versions of the VGG16 [90] and ResNet[96] architectures with transfer learning, data augmentation and some pre-processing work.

Similarly, Yu *et al.*[163] , Kwasigroch *et al.*[164], Lopez *et al.*[165], Kassani *et al.*[166] presented studies on cancerous dermoscopic skin lesions classifications using DCNNs with transfer learning. Some of these works are done using InceptionNet-v3[95] and VGG-Net[90].

Shorten *et al.*[157] discussed a few limitations on data augmentations such as:

- Finding the optimum final post-augmented dataset size to produce the best performing model. There is a possibility that the augmented dataset can be heavily biased.
- There are no existing augmentation techniques that can correct a training dataset with very poor diversity with respect to the testing data. All the augmentation algorithms perform best under the assumption that the training data and testing data are both drawn from the same distribution. Hence, these limitations in data augmentation could be a potential problem for small medical image datasets because of class imbalance and diversity.

Morid *et al.*[167] systematically reviewed the literature on approaches to transfer learning in medical image analysis that are based on CNN models trained on the non-medical ImageNet dataset for medical image analysis. A vital research gap discussed in this review is finding the optimal dataset size that can support medical image analysis tasks, as a large dataset may not always be available.

An extensive survey published by Pan *et al.*[158] discusses a few limitations of transfer learning that may apply to medical image analysis, such as negative transfer, which is an open problem in transfer learning. For example, when using

the ImageNet[3] dataset for medical/clinical image analysis, there is no similarity between the source and target domains. ImageNet contains real-world objects, animals, fruits, balloons etc. Hence there is a high possibility of performance intrusion in the target domain; which is known as negative transfer [168]. Negative transfer is made more likely by the fact that well performing CNN models are pre-trained on a non-medical dataset [167]. In simple words, it is not yet established,

- Which characteristics facilitate an effective transfer of features and weights in the transfer learning and fine-tuning process?
- Whether the features that are transferred from ImageNet are plausible or not?
- Whether the transferred features and weights are plausible, how we can quantify that?
- At what level is it adequate to incorporate the features of a non-medical dataset during the training process?

Table 2.4: Comprehensive Summary of Research on Data Augmentation and Transfer Learning in Skin Disease Classification. This table delineates a curated list of studies that have leveraged both data augmentation and transfer learning techniques to enhance the accuracy and reliability of skin disease classification models. It details the authors, publication year, specific skin conditions targeted, types of data augmentation and transfer learning methods used, datasets employed, and the performance metrics achieved. The table further illustrates how the combination of data augmentation and transfer learning contributes to overcoming challenges such as data scarcity and model generalizability in dermatological imaging, reflecting the synergy between these two methods in advancing the field of medical diagnostics.

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source	Dataset volume in total/per class	Methodology	Best results and Per- formance measures
Esteva <i>et al.</i> [82], 2017	2,032 skin diseases for Training the model and Tested for malignant melanomas, benign nevi, malignant basal, squamous cell carcinomas, intraepithelial carcinomas, pre-malignant actinic keratosis, benign seborrheic keratosis.	ISIC Dermoscopic Archive, Edinburgh Dermofit Library and data from the Stanford Hospital.	1,29,450 images. 2,032 disease classes for Training and 7 types of cancerous lesion classes for Testing.	Pre-trained on ImageNet dataset Transfer Learning, Data Augmentation, InceptionNet-v3.	Accuracy =93.33%, Confusion matrix, Saliency Maps, Sensitivity-specificity curves.

Continuation of Table 2.4

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source	Dataset volume in total/per class	Methodology	Best results and Per- formance measures
Sourav Mishra <i>et al.</i> [169], 2018	9 common skin conditions: Acne, Alopecia, Crust, Erythema, Leukoderma, Pigmented Maculae, Pustule Ulcers and Wheal.	N/A	Each class comprises of approximately 4600 images in which the division between training and test ratio of 90:10	Pre-trained on ImageNet DC-NNs such as: ResNet18, ResNet50, ResNet152, DenseNet161.	Classification accuracy: 82.30%(by ResNet152)NVIDIA Titan XP and CUDA v8
Binol <i>et al.</i> (RosNet) [99], 2019	Rosacea lesions	Ohio State University (OSU) Division of Dermatology (using DSLR camera)	41 facial images. The size of each image is 4608×3072	Pre-trained on ImageNet DC-NNs: Inception-ResNet-v2, ResNet-101, Image classification problem for rosacea and non-rosacea lesions, Data Augmentation, Anatomically directed post-processing (anthropometric model)	Dice co-efficient: 92.9% False positive rate, MATLAB R2018b using the Deep Learning Toolbox, (HPC) with 128 GB RAM and 16 GB NVIDIA Tesla P100 PCI-E GPU.
Goceri [160], 2019	5 common skin diseases; (1) Acne vulgaris, (2) Hemangioma, (3) Psoriasis, (4) rosacea, and (5) Seborrheic dermatitis.	N/A	Total = 800, Per class= 160	Pre-trained on ImageNet: U-net, InceptionNetV3, InceptionResNetV2, VGGNet and ResNet.	Classification Accuracy: 80% (by ResNet50). GeForce GTX 980Ti GPU, Intel Core i7-4930 K processor, 6GB memory and 16GB RAM
Sun <i>et al.</i> [126], 2016	198 Common skin diseases: eczema, psoriasis, acnevulgaris, pruritus, alopecia areata, decubitus ulcer, urticaria, scabies, impetigo, abscess, bacterial skin diseases, viral warts, molluscum, melanoma and non-melanoma skin cancer	SD-198 and SD-128	total= 6,584; for some classes-10 to 20 samples.	Pre-trained on ImageNet DC-NNs such as: CaffeNet, CaffeNet+ finetuning, VGG Net, VGG Net + finetuning	Classification accuracy: 50.27% (by VGGNet + finetuning)
Yang <i>et al.</i> [170], 2018	198 common skin diseases.	SD-198	6,584; for some classes-10 to 20 samples.	Pre-trained on ImageNet DC-NNs such as: GoogleNet, GoogleNet + fine tuning, ResNet, ResNet + fine tuning	Classification accuracy: 53.35 % (by ResNet + fine tuning)

Continuation of Table 2.4

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source	Dataset volume in total/per class	Methodology	Best results and Per- formance measures
MAA[161] 2019	Seven skin diseases- Melanoma (1113), Melanocytic nevus (6705), Basal cell car- cinoma (514), Actinic keratosis (327), Be- nign keratosis (1099), Dermatofibroma (115) and Vascular (142).	ISIC 2018 Melanoma Detection Challe- nege and Dataset	Training set = 10015 skin lesion images. The validation dataset = 193 skin lesion images.	Data Augmentation. Rep- resentation learning. Pre- trained on ImageNet DCNNs with fine tuning such as: PNASNet-5-Large, Inception- ResNetV2, SENet154, Incep- tionV4, An Ensemble of all models.	Validation score: 76% (by PNASNet-5-Large)
W.Sae- Lim <i>et</i> <i>al.</i> [171], 2019	Seven skin diseases: Cancerous	Human Against Machine 10,000 (HAM10,000)	10,015 images	Pre-trained on ImageNet Modified Mobile-Net with Data Augmentation Data up-sampling.	Accuracy: 83.23%, Specificity: 87%, Sensitivity: 85%, F1 score:82%
Kemal <i>et al.</i> [172], 2020	Seven skin diseases: Cancerous	HAM10,000	10,015 images	A CNN architecture + One verses all which 1,243,463 pa- rameters in total. Data Aug- mentation.	Average precision: 92.90%
Hosny KM <i>et</i> <i>al.</i> [173], 2019	Melanoma skin lesions.	2017 ISIC challenge dataset, MED- NODE, DermIS+ Der- mQuest.	2000, 170, 206	Data Augmentation Pre- trained on ImageNet for transfer learning on AlexNet	Average Accuracy: 95.91%, 96.68%, 97.07%.
Mahbod <i>et</i> <i>al.</i> [174], 2019	411 malignant melanoma (MM), 254 seborrheic ker- atosis (SK) and 1372 benign nevi (BN)	ISIC 2016, ISIC 2017	Total= 2037, Training= 1887, Valid- ation set=150	multi-class non-linear support vector machine (SVM) clas- sifiers. Fusion of DCNNs such as AlexNet+ VGG16+ ResNet18)	Average AUC:90.69%
Mendes <i>et</i> <i>al.</i> [175], 2018	11 distinct lesions with 4 malignant illness.	MED- NODE, Edinburgh Dermofit library, Atlas	170, 1300, 3816	Data Augmentation, Pre- trained on ImageNet DCNNs: ResNet-152	Total accuracy: 78% (by ResNet-152)
Cui <i>et</i> <i>al.</i> [162], 2019	Melanoma (295) and non-melanomas (311)	International Society for Digital Imaging of the Skin (ISIC).	Total= 606	Image pre-processing seg- mentations, feature extrac- tions, Machine Learning classifications- using SVM, Regression tree, K-nearest neighbour, Logistic regres- sion, Transfer Learning using-AlexNet, VGG16, VGG19, Google Inception v3.	Average accuracy, Sensitivity, Specificity. (Accuracy, Sensitivity, Specificity=93.70%, 95.30%, 92.10% re- spectively by Inception v3), Windows 7 system, MATLAB R2018b, TensorFlow 1.3, GTX1080Ti (Nvidia).

Continuation of Table 2.4

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source	Dataset volume in total/per class	Methodology	Best results and Per- formance measures
Liu <i>et al.</i> [159], 2020	26 most common skin conditions in adult, by making it 419 categories of skin conditions (Rosacea is not included), Labeled by a cohort of 37 US board-certified and 5 Indian board-certified dermatologists.	Tele Dermatology consultation dataset (Confidential)	Training=64,837, Validation=11,268, clinical meta-data (demographic information and medical history)	Transfer Learning pretrained on ImageNet, Fine tuning, Inception v4, Data Augmentation	top-3 accuracy=93% - top-k sensitivity=83% -95% confidence intervals, -validated by 18 board certified dermatologists who did not participate in labelling the input images.
End of Table					

2.5.2 Generating synthetic images using GANs

As discussed in the previous Section 2.5.1., data augmentation techniques are used in various studies, but there are a few notable limitations. Although data augmentation techniques help in transforming the images by zooming, cropping, flipping, rotating, it does not radically improve results when there is only a handful of data available for some specific skin conditions. However, GANs can be explored in creating synthetic data from an existing limited dataset without splitting the dataset into training, validation, and test sets.

Generative models are inspired by the unsupervised learning model approach. They can generate new examples that are similar to the training images. The GAN framework [40] consists of a pair of adversarial networks – a Generator Network G and Discriminator Network D . The Generator Network G tries to transform random noise from the prior distribution over the input variables (usually a standard normal distribution/gaussian distribution) to generate fake/synthetic images which look as realistic as possible. The input variables to G are drawn from a normal distribution and the output is a synthetic image. Generally, the dimension of the output image is much greater than the dimension of the input variables.

Simultaneously, a Discriminator Network D attempts to discriminate between the sample images obtained from the real training data and the fake/synthetic images obtained from the generator function G . By utilizing the feedback from the discriminator D , the parameters of the generator G can be adjusted such that its samples are more likely to fool the discriminator network in its classification task. Ultimately, it is desired that the distribution of the fake images has as much in com-

mon as possible with the real images. In the Discriminator Network D , the input is an image, and the output is a real number between 0 and 1, which represents the probability that the input image is real. Ideally if D is working properly, the output will be close to 1 for a real image and close to 0 for a synthetic one.

The equation below represents the function V which GANs optimise during training. x_{real} represents a real image. z represents the random input values for G . In this equation, the first term only applies to real data and the second term only applies to the synthetic data.

$\mathbb{E}_{x \sim p_{data}(x)}$ indicates the expected value of $\log(D(x_{real}))$ where $p_{data}(x)$ is the probability distribution over the real data and $\mathbb{E}_{z \sim p_z(z)}$ is the expected value of $\log(1 - D(G(z)))$ where $p_z(z)$ is the probability distribution over the input values to G . The parameters of G and D are optimized by playing a minimax game, which involves varying the parameters of G to minimise V in an outer loop and varying the parameters of D to maximise V in an inner loop. The value function $V(G, D)$ is defined as:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x_{real}))] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (2.1)$$

At the early stages of the learning process, G will not be generating any realistic looking images and D can reject samples with high confidence because they are clearly different from the training data. In this case, $\log(1 - D(G(z)))$ should be close to zero because $D(G(z))$ will be close to zero [143, 40].

In skin disease classification and analysis, there have been a few impactful studies, which support generating synthetic images from existing datasets of real images to increase the training samples and improve the classification accuracy. These studies will be considered as a part of a critical analysis and are listed in Table 2.5.

The main points that can be drawn from Table 2.5 are:

- Most of the work on GANs for skin diseases classification is from 2018 onwards.
- Most of the diseases considered for generating the synthetic image dataset are melanoma or cancerous skin lesions. There is no work related to rosacea or any facial skin conditions.
- Minimum of 2,000 input real images are considered for generating synthetic images.
- Majority of the works use the ISIC 2017 and 2018 datasets.

- DCGAN, PGAN and LAPGAN appear to be popular architectures used in these studies.

Qin *et al.* [176] states that the generation of synthetic samples for vascular lesions (142 input images) makes the lesion region look realistic. However, the skin texture around the lesion is still fuzzy and lacks contrast in some samples. The same effect appears in melanoma images. For melanocytic nevus (6705 images), the most representative images are with concentrated colour, clear edge and regular shape. Hence, more training samples lead to generating better quality synthetic ones. In Rashid *et al.* [177], the highest F1 score was obtained in the melanocytic nevus (6705 images) category because of the high number of samples available.

Major takeaways on DCGAN, LAPGAN and PGAN

DCGAN [178] and LAPGAN [179] have proven to work well for generating synthetic images from the input noise, albeit at lower resolutions such as 64×64 pixels. While these methods have been foundational, they are typically characterized by limitations such as checkerboard artifacts in DCGAN and high-frequency artifacts in LAPGAN images, as discussed by Baur *et al.* [180, 181]. These limitations underscore the need for improved resolution and texture quality in synthetic image generation.

Conditional GAN approaches [182] have been suggested to improve upon these limitations by providing additional, relevant information to both the generator and discriminator, especially useful in cases of imbalanced datasets. On the other hand, Progressive Growing of GANs (PGAN) [183] has made significant strides in realistic image synthesis, achieving high resolutions up to 1024×1024 pixels, and showing promise in generating highly realistic images without the need for conditioning. Despite these advancements, challenges such as the accurate rendering of complex features like hair in medical images remain, indicating the need for further research and methodological enhancements.

Baur *et al.* [180, 181] also highlight several research gaps critical to advancing the field of synthetic image generation using GANs:

- Whether there is an information gain in the synthetic samples over the actual training dataset?
- Whether the gain is higher than using conventional data augmentation?

- How many training images are required to obtain reliable generative models?
- Whether there is still a need to enhance the methodology to account for filamentary structures?

Table 2.5: Detailed Analysis of Studies Utilizing Generative Adversarial Networks (GANs) in Skin Disease Classification. This table provides an in-depth overview of significant research where GANs have been applied to the field of dermatological imaging. It includes essential details such as the authors, publication year, specific skin diseases studied, names and sources of datasets used, dataset volumes both in total and per class, and the methodologies or benchmarks employed in each study. Additionally, the table highlights the best results achieved in terms of evaluation metrics, the number of samples generated, deployment strategies, and hardware used. This comprehensive summary offers insights into how GANs have been effectively utilized to generate synthetic data for skin disease classification.

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results Evaluation Metrics /Samples Gen- erated, Deployment and H/W
Qin <i>et al.</i> [176], 2020	Seven skin diseases- Melanoma (1113), Melanocytic nevus (6705), Basal cell car- cinoma (514), Actinic keratosis (327), Be- nign keratosis (1099), Dermatofibroma (115) and Vascular (142)	International Skin Imag- ing Col- laboration (ISIC) 2018	10,015 der- moscopic images. 600×400 pixels and 96dpi.	Transfer learning and fine tuning on Transfer-ResNet50, GAN, DCGAN, StyleGAN, SL-StyleGAN(proposed work), Non-linear mapping network, Adaptive Instance Normaliza- tion (AdaIN)operations, Style Mixing, Stochastic variation	Inception Score, Fréchet Inception Dis- tance (FID), Precision and Recall. Best result obtained by SL- StyleGAN proposed method. H/W: Intel Xeon Gold 6144 with 192 GB RAM, GPU of NVIDIA Quadro P40 0 0.
Lei <i>et al.</i> [184], 2020	Melanoma skin lesions	International Skin Imag- ing Col- laboration (ISIC) Skin Lesion Challenge Datasets 2016, 2017 and 2018	ISIC2016: Training- 900, Test-379; ISIC2017: Training- 2000, Testing- 600; ISIC2018: Training- 2296, Testing- 300	Segmentation task, A deep encoder-decoder module UNet- SCDC (skip connection and dilated convolution) (proposed work), Dual discrimination module.	Accuracy, Sensitivity, Specificity, Jaccard In- dex, Dice coefficient, H/W: Two NVIDIA TITAN XP GPUs.

Continuation of Table 2.5

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results Evaluation Metrics /Samples Gen- erated, Deployment and H/W
Baur <i>et al.</i> [180], 2018.	Benign and malignant skin cancerous lesions	ISIC2017	2000 (256×256px)	DCGAN, LAPGAN, deeply discriminated GAN (DDGAN) (proposed work), Transfer learning; pretrained ResNet-50	Earth mover's distance (MD) (Wasserstein-Distance), JS Divergence, 2000 random samples generated. H/W: NVidia 1080Ti
Rashid <i>et al.</i> [177], 2019	Melanoma (MEL); Melanocytic Nevus (NV); Basal Cell Carcinoma (BCC); Actinic Keratosis (AKIEC); Benign Keratosis (BKL); Dermatofibroma (DF); Vascular Lesion (VASC)	ISIC2018	8,000 training images, 2,000 testing images.	Transfer learning, Finetuning, DenseNet, ResNet50, GAN based augmentation	Precision and Recall, F1 score, Balance accuracy score: 0.86 using GAN based augmentation, Highest F1 score for Melanocytic Nevus (NV).
Baur <i>et al.</i> [181], 2018	Benign and malignant skin lesions of 7 categories.	ISIC2018	10,000 labelled training samples	PGAN, DCGAN, LAPGAN, Visual Turing Test	Sliced Wasserstein Distance (SWD): 20.0197 (closest to the lower bound, 10,000 synthetic images generation per model, User study among 3 expert dermatologists and 5 Deep Learning experts; showing experts had a hard time distinguishing real and fake image
Bisla <i>et al.</i> [185], 2019	Three classes: melanoma, nevus, and seborrheic keratosis.	ISIC 2017 (3 classes) PH2 (2 classes), Edinburgh Dataset, Test dataset: ISIC 2017 and ISIC 2018	803+40+76 cases of melanoma, 2107+ 80+ 331 cases of nevus, and 288+ 257 cases of seborrheic Keratosis from ISIC2017, PH2 and Edinburgh Dataset respectively	Segmentation using U-Net architecture, de-coupled DCGANs for data generation, Pre-trained ResNet-50 for final classification.	MSE for GANs, 350 synthetic images for melanoma and 750 synthetic images for seborrheic keratosis (26% artificially generated data for training) ROC and AUC for classification.

Continuation of Table 2.5

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results Evaluation Metrics /Samples Gen- erated, Deployment and H/W
Bissoto <i>et al.</i> [186], 2019	Melanoma skin lesions	ISIC 2017 Challenge, ISIC Archive, Dermofit Image Library, PH2 Dataset, For Test: Interac- tive Atlas of Der- moscopy	2,000, 13,000, 1,300, 200, 900 der- moscopic images of 270 melanomas categories.	SLIC algorithm, DCGAN, Conditional PGAN, pix2pixHD GAN (a conditional image- to-image translation GAN) using only semantic map, Real+Instance+PGAN	AUC, p-value
Pollastri <i>et al.</i> [187], 2019	Melanoma Skin lesions	ISIC 2017	1882 (As Authors wished to remove 118 images from the dataset	Segmentation task using Base- line CNN architecture and U- Net, DCGAN, LAPGAN	Jaccard Index
Ghorbani <i>et al.</i> [44], 2020	26 skin conditions (no rosacea) Melanocytic nevus, Melanoma and Seborrheic Keratosis/Irritated Sebor- rheic Keratosis, Scar condition, Basel cell Carcinoma etc.	Tele- dermatology service dataset: collected in 17 clin- ical sites in two U.S. states from 2010 to 2018	9,897 cases and 49,920 images; each case contains one or more high resolu- tion images (resolu- tion range: 600×800 to 960×1280).	Pix2pix GAN architecture, Hu- man Turing Test, DermGAN, MobileNet	20,000 synthetic im- ages are generated using the 8-class Derm- GAN model and added them to the existing training Data to train MobileNet.
Romsaas <i>et al.</i> [188], 2020	Benign and Malignant cancer lesions	ISIC 2019	Total= 25,331 im- ages. With 7 classes.	ACGAN (128×128), CycleGAN (256×256), Path-Rank-Filter (to generate class specific synthetic images)	Upto 24,000 synthetic images are generated using ACGAN and CycleGAN, Preci- sion/recall, Maximum Accuracy = 86%
Bissoto <i>et al.</i> [189], 2021	Benign and malignant cancer lesions	ISIC 2019, ISIC 2020, Derm-7pt- dermoscopic, Derm-7pt- clinical, Dermofit.	3,863, 1,743, 872, 839, 973	PGAN and StyleGAN2 (to generate from random noise), pix2pixHD and SPADE (for semantic segmentation masks to guide the generation), Inception-v4 for classification, Data Augmentation.	Fréchet inception distance (FID) "Style- GAN2 performed ahead of all other GANs".
End of Table					

Style-based Generative Adversarial Networks

To examine a specific facial skin condition such as rosacea, it is necessary to access full face images with anatomical details. Therefore, a GAN model should be trained to generate full facial synthetic images rather than partial facial images. The Style-based GAN [190] is an improvement of PGAN, which supports generating higher-quality images via the incremental expansion of both generator and discriminator models for low quality to high quality images. Looking at the state of the art, Bissoto *et al.*[189] show promising results with StyleGAN2 [43].

Noise-based GANs, such as StyleGAN2 [43], use a random noise vector as input to generate new images from scratch. This approach is particularly effective for creating diverse and detailed images that do not need to correspond to any specific input image, relying on the model’s learned distribution of the data.

In contrast, translation-based GANs, exemplified by Pix2pixHD [191] and SPADE [192], focus on translating one possible representation of an image to another, such as converting a segmentation map to a photorealistic image. These models are typically conditioned on an input image, making them suitable for tasks where a direct correspondence between the input and output images is necessary, such as photo editing or image-to-image translation. Bissoto *et al.*’s [189] study suggests that noise-based GANs work better than translation based GANs.

Chai *et al.*[193] looked at one of the open questions on GANs which aims to look at how to convert unstructured latent code to a high quality output while maintaining global consistency. They show that StyleGAN [190] performs better in separating faces from background artifacts. Subsequently, StyleGAN2 [43] was introduced as an improved version of StyleGAN, hence StyleGAN2 can be considered for obtaining relatively high-quality synthetic facial images with global consistency of anatomical details and affected regions of the face. StyleGAN2 works better in comparison to PGAN. However, StyleGAN2 takes up to one month of GPU time for a single course of training. Additionally, StyleGAN2 also produces high quality facial images [43], which can be useful for generating synthetic face images with facial diseases. Recently, StyleGAN2 with adaptive discriminator augmentation (StyleGAN2-ada) [194] has been introduced, which claims to help with relatively limited data (a few thousand training images) regime due to an adaptive discriminator augmentation mechanism which does not require changes to the network architecture or defined loss functions. This may help in generating images with a few hundred samples. As StyleGAN2-ada has acquired optimal results with Flickr Faces-HQ dataset [190],

which is basically a facial dataset, it may help in generating synthetic images of facial skin conditions e.g. rosacea.

2.5.3 Meta-Learning and Few-Shot Classification

The concept of meta-learning and few-shot classification can help deal with the limited amount of data without any transfer learning techniques, and without any augmentation process (such as synthetic image generation) of the dataset, but through hyperparameter optimization.

Meta-learning, often described as ‘learning to learn,’ involves designing models that can quickly adapt to new tasks with minimal data by leveraging prior knowledge and learning experiences. This approach optimizes the learning process itself through hyperparameter optimization and model architectures that are inherently flexible. Meta-learning enables the model to generalize from a small number of examples in a new context, a process that is crucial for tasks with limited data availability.

Few-shot classification, on the other hand, is a specific application of meta-learning where the model is designed to learn from only a few examples—often as little as one or a handful—per class. The challenge in few-shot classification is to make accurate predictions in settings where traditional models would struggle due to the scarcity of data.

While both meta-learning and few-shot classification are related in their aim to make efficient use of small datasets, they are not the same. Meta-learning encompasses a broader set of techniques for improving the learning algorithms, which can include, but is not limited to, few-shot classification scenarios. Few-shot classification is a specific problem setup that meta-learning techniques can address.

In a meta-learning problem, we have a meta-training set and a meta-test set, each of which contains a number of “tasks”. Each task is associated with a training set and a test set containing both feature vectors and correct labels. It is similar to a standard machine learning problem but each task is considered as one data sample [146]. The goal of meta-learning is to acquire generic knowledge of different tasks. The knowledge can then be transferred to the base level learning to provide generalisation in the context of a single task [195].

Meta-learning assumes that all the tasks in both the meta-training set and the meta-test set, have some degree of similarity. For example, in the case of skin diseases the meta-training tasks might represent different, well-studied skin conditions and

the meta-test set might represent a rarer skin condition, for which there is limited data.

Regular machine learning involves the minimisation of a loss function \mathcal{L} .

$$\theta_C^* = \arg \min_{\theta} \mathcal{L}(\theta_C, D^{tr}) \quad (2.2)$$

where θ_C is a vector of the parameters of a classifier (e.g. the weights of a neural network). The classifier takes an input x which represents the feature values of an example and outputs a label y .

Meta-learning involves estimating the parameters θ of a “learning function” f_{θ} . The learning function takes as input a training set D_i^{tr} and outputs a vector ϕ_i as in Equation. 2.3 below.

$$\phi_i = f_{\theta}(D_i^{tr}) \quad (2.3)$$

ϕ_i can be used to generate a classifier, which can classify examples from the corresponding test set D_i^{ts} for task i .

Meta-learning involves minimising the function in Equation. 2.4 with respect to θ ,

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \mathcal{L}(\phi_i, D_i^{ts}) \quad (2.4)$$

where n is the number of tasks in the meta-training set. For each task i a vector ϕ_i is generated using the learning function f_{θ} as in Equation. 2.3 above. The same learning function (with the same values for θ) is used for all tasks.

Then ϕ_i is used to generate a classifier, which can then applied to the test set D_i^{ts} to generate labels for each example in the test set. The loss function \mathcal{L} is calculated by comparing the generated labels with the true labels. The loss functions for all the tasks in the meta-training set are then summed. An optimisation algorithm is then used to find the values of θ which minimises the function in Equation. 2.4.

Fig. 2.2 is an example of a meta-learning setup. The figure represents the meta-training set and the meta-test set, where each grey rectangle is a separate task that consists of a training set and a test set separated by the dotted lines. The training sets are also called support sets. The test sets are also called query sets. Each image is taken as one example within the dataset. Each training/support set has 5 different example images, and each test/query set has 2 additional example images. The meta-learning model can be trained on the tasks in meta-training set and then tested on entirely new tasks in the meta-test set. The whole dataset is divided and structured as shots/iteration/episodes.

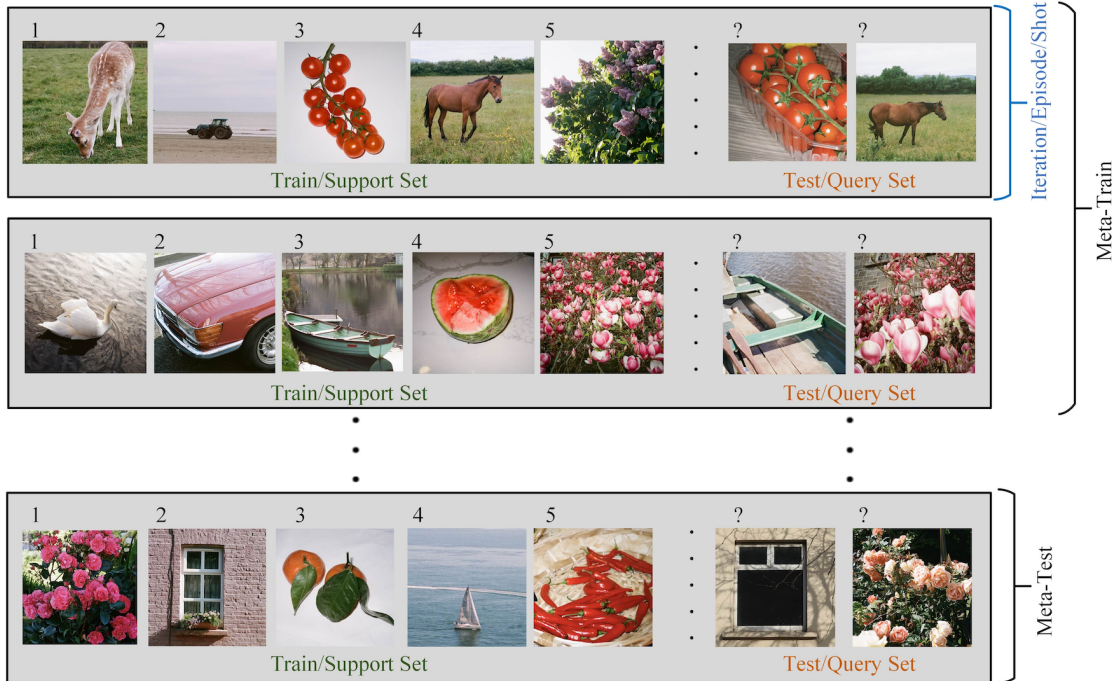


Figure 2.2: An example of Meta-Learning set up in 1-shot-5-way classification.

Few-shot classification is a subset or specific application of meta-learning. “Few-shot learning” means learning a large number of tasks using datasets containing only a few examples known as ‘shots’, e.g. training five different classes where each class has just one example is called 1-shot-5-way classification. Fig. 2.2 is an example of a 1-shot-5-way classification task (the word ‘way’ indicates a class). However, the ‘5-way’ aspect is not a fixed requirement; it simply indicates the number of classes involved in the classification task. The model can be configured for ‘N-way’ classification, where ‘N’ can be any number representing the diversity of classes to be identified, such as ‘1-shot-10-way’ or ‘5-shot-20-way’.

The main idea of few-shot meta-learning is to train a few iterations/shots from the whole dataset, so the model can quickly adapt to the new task with only fewer examples. In order to attain this idea, the meta-learner is trained over a large set of different tasks, such that for the new unseen tasks, the model can learn quickly with only a limited number of examples. In effect, the meta-learning problem treats an entire dataset as training examples [147, 148]. The whole dataset is divided and structured as shots/iteration/episodes.

There are three general meta-learning categories including: Model-based (Black-Box) Meta-Learning, Metric-based (Non-parametric) Meta Learning, Gradient-based (Optimization-based) Meta Learning. Typical architectures that are model-based meta learning are Memory Augmented Neural Network (MTNN) [196], Meta-Net

[195], Simple Neural AttentIve Learner (SNAIL) [197]. Typical architectures that are metric-based meta learning are: MatchingNets [198], RelationNet [199], FSL with Graph Neural Networks [200], Prototypical Networks [201]. Typical architectures that are gradient-based meta learning are: MAML [148], Reptile [202], Auto-Meta [203], GIMLI [204], ALFA [205]. The Model-based meta-learning is conceptually very simple, but it has minimal inductive bias, meaning everything must be meta-learned. However, metric-based meta-learning can work very effectively by combining some inductive bias with easy end-to-end optimization and is restricted to classification models. On the other hand, optimization-based meta-learning is convenient to apply any architecture and good at generalizing a wide range of domains. An overview of the state-of-the-art studies on skin disease analysis using meta-learning and few-shot learning is presented in Table 2.6.

Since meta-learners often face a wide variety of tasks during training, augmentations can ensure that the model does not learn to overfit the idiosyncrasies of the training tasks and instead develops a more generalized learning strategy. A few other studies on meta learning and few-shot learning for other domains of medical image analysis shown promising results. Zhao *et al.*[206] proposed a novel approach for automating data augmentation for synthesizing labelled MRI brain scans. The proposed method only requires a single segmented scan. This approach uses a semi-supervised method for dealing with unlabelled scans that results in a major improvement over bio-medical image segmentation state-of-the-art methods.

Guo *et al.*[207] and Zhao *et al.*[109] showed a comparative study among different CNN architectures using meta-learning and few-shot learning algorithms in various medical image analyses. Patacchiola *et al.*[208] proposed Deep Kernel Transfer (DKT) which is a Bayesian treatment for the inner loop through deep kernels in the meta-learning algorithm. This approach has many advantages such as, (a) it is simple to implement as a single optimizer, (b) it offers uncertainty quantification, and (c) it does not need estimation of task-specific parameters.

Cai *et al.*[209] proposed a novel score-based meta transfer-learning to address the cross-domain few-shot learning problem. This work claims to achieve an average accuracy of 74.06%, which significantly outperforms previous best-performing meta-learning and transfer-learning methods by 14.28% and 5.93%. Sun *et al.*[210] have proposed a novel few-shot learning method called Meta-Transfer Learning (MTL) which learns to adapt a DNN for few-shot learning tasks.

The datasets used in the study by Zhao *et al.* [109] are miniImageNet (100

classes with 600 samples per class) and Fewshot-CIFAR100 (100 classes with 600 samples per class). Compared to other few-shot learning methods, MTL has shown remarkable results in low data settings. In a few-shot learning pipeline, Gidaris *et al.*[211] used self-supervision as an auxiliary task, that allows feature extractors to learn richer and more transferable visual representations while using few annotated samples. Hospedales *et al.* [212] reviewed many existing works in various domains. They claim that there is still a gap in understanding which kinds of meta-representations tend to generalize better under certain types of domain shifts. Wang *et al.*[213] looked at various studies on a few-shot learning approach in which it is said that real-world computer vision tasks with low data situations are the first test-bed for few-shot learning algorithms. It is also seen in various works that transfer learning methods are used in a few-shot learning approach as part of the domain adaption, which might be beneficial in case of limited data availability across the medical image analysis domain.

Table 2.6: Comprehensive Review of Studies Utilizing Meta Learning and Few-Shot Classification in Skin Disease Diagnosis. This table collates a selection of significant research efforts that have applied meta-learning and few-shot classification strategies to tackle skin disease diagnosis challenges. It catalogs the studies by author, publication year, the specific skin diseases investigated, and the datasets utilized, including the names, sources, and volumes both in total and per class. Moreover, it details the methodologies and benchmarks used in each study along with the best results achieved, outlining the evaluation metrics, as well as deployment strategies and hardware requirements.

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results/ Evaluation Metrics/ Deployment and H/W
Li <i>et al.</i> [214], 2020	Melanocytic nevus (6705), melanoma (1113), benign keratosis (1099), basal cell carcinoma (514), actinic keratosis (327), and dermatofibroma (115). squamous cell carcinoma, haemangioma, and pyogenic granuloma	ISIC2018 Skin Lesion Dataset, Dermofit Image Li- brary (for validation)	10,015 First four dis- eases for training and the remaining 3 are for testing	Difficulty aware meta learning (DAML) framework, Transfer learning.	AUC: 83.3% (for 5 samples per class).

Continuation of Table 2.6

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results Evaluation Metrics/ Deploy- ment and H/W
K. Mahajan <i>et al.</i> [215], 2020	Dermoscopy images eczema, acne, and various cancerous conditions.	ISIC 2018, Derm7pt, SD-198	10,015 (7 classes), 2,000 (20 classes), 198	Reptile (Gradient based meta learning), Prototypical networks (Distance metric based meta learning technique which computes a prototype vector as the representation of each class), G-convolutions (group equivalent convolutions), 2-way classification.	For ISIC2018: AUC: 86.8 (using Reptile, G-conv, 5-shot) For Derm7pt:AUC: 77.2 (using Reptile, G-conv, 5-shot) For SD-198:AUC: 89.5 (using Reptile, 2-way G-conv, 5-shot)
Zhang <i>et al.</i> [216], 2020	Melanocytic nevus (6705), melanoma (1113), benign keratosis (1099), basal cell carcinoma (514), actinic keratosis (327), vascular lesion (142) and dermatofibroma (115)	ISIC2018	10,015(7 classes) 4 classes with the largest number of samples for training and the remaining 3 classes with relatively fewer number of samples for testing.	Fine tuning and data augmentation, MAML (Model Agnostic Meta Learning), ST-Meta Diagnosis Network, A few shot settings (3ways1shot, 3ways3shot and 3ways5shot) with different STN (Spatial Transform Network) modules stacked up into different layers.	Average Accuracy: 81.38%.
Fayjie <i>et al.</i> [108], 2020	Miscellaneous	FSS-1000 (Base Training set), ISIC 2018, PH2 dataset	1000 classes where each class contains 10 images, 2594 dermoscopic images with their respective masks 200 RGB dermoscopic images of melanocytic lesions.	Segmentation task, Pretrained VGG network on ImageNet is used as Encoder, A few shot learning with ‘support’ and ‘query’ set, The pretrained VGG network architecture from ImageNet is used as an encoder, Atrous convolutions with dilation rate 2 are incorporated with episodic training.	DSC (Sørensen–Dice coefficient)score, The proposed approach outperforms the few-shot baseline by a margin between 6-7%, H/W: Keras with Tensorflow using the Nvidia Titan X GPU.

Continuation of Table 2.6

Author/ Index/ Year	Skin Disease Names	Dataset Name/ Source/	Dataset volume in total/per class	Methodology/Benchmark	Best results Evaluation Metrics/ Deploy- ment and H/W
Zhao <i>et al.</i> [109], 2020	Miscellaneous (no rosacea)	miniImagenet (source domain) Crop Diseases dataset, EuroSAT, ChestX and ISIC 2018 (target domain datasets)	N/A	Few-shot learning, Baseline fine tuning method, BSR (Batch Spectral Regularization) with data blending, BSR with LP (Label Propagation), BSR with data blending in Ensemble, BSR with data blending and with LP in the Ensemble network	Best-Average Per- formance results are obtained by BSR with data blending and with LP in Ensemble network architecture using 5-way 50-shot.
End of Table					

2.5.4 Federated Learning

Building upon the critical discussion on data scarcity, transfer learning, GANs, meta-learning, and few-shot classification, in this section we delve into Federated Learning (FL) [217] as a synergistic approach. FL enhances the collective intelligence in skin disease diagnosis by enabling a decentralized, collaborative model training that respects data privacy and overcomes the inherent limitations of small datasets, promising a significant leap forward in medical imaging and in machine learning driven skin disease classification. We explore the application of FL in the context of skin disease analysis, its potential in enhancing the diagnosis accuracy with limited datasets, and how it aligns with the current technological advancements and challenges outlined in the field.

Federated learning (FL) is a machine learning approach where multiple parties collaboratively train a model while keeping their individual data localized and private. Instead of sharing data, participants share model updates, which are then aggregated to improve a central model. This method addresses data privacy, security, and access issues, allowing for robust model development without compromising sensitive information.

Federated Learning works by allowing multiple devices or servers (participants) to collaboratively train a machine learning model while keeping the data localized. Each participant trains a shared model using their own data, calculates updates (like

gradients), and sends these updates to a central server or aggregator. The aggregator combines these updates to improve the model and then sends the updated model back to participants. This process iterates, improving the model with each round, without ever sharing the raw data itself, thus preserving privacy and reducing data transfer needs.

In Federated Learning, where the goal is to minimize a global loss function, representing the model’s performance across all clients, without sharing the local data. The loss is calculated as a weighted sum of local losses computed on each client’s data [218]. This ensures that learning is collaborative and privacy-preserving.

$$\min_{\phi} \mathcal{L}(X; \phi) \text{ with } \mathcal{L}(X; \phi) = \sum_{k=1}^K w_k \mathcal{L}_k(X_k; \phi) \quad (2.5)$$

Where,

- $\mathcal{L}(X; \phi)$: The global loss function measuring the model’s performance across all clients.
- K : The number of local clients participating in the federated learning process.
- \mathcal{L}_k : The local loss function for the k -th client, computed using their private data X_k .
- w_k : Weight coefficients reflecting the contribution of the k -th client’s local loss to the global loss function.
- \min_{ϕ} : Represents the optimization goal to find the model parameters ϕ that minimize the global loss function.

There are commonly used strategies and algorithms used for federated learning. These include Centralized Federated Learning, Decentralized Federated Learning, and Heterogeneous Federated Learning (HeteroFL), each offering unique approaches to model training and collaboration among diverse data sources. Understanding these types provides a comprehensive overview of how federated learning can be adapted and implemented for various scenarios, particularly in handling limited datasets in skin diseases.

Centralized Federated Learning: In centralized federated learning, a central server is the linchpin of the network, coordinating the training process across different clients. Each client device, or node, trains a model on its local data and then sends only the model updates—not the data itself—to the central server. The server

aggregates these updates to improve the global model, which is then sent back to the clients for further training. This cycle continues until the model is adequately trained. While this method can produce highly accurate models due to the centralized aggregation, it is susceptible to bottlenecks and potential points of failure if the central server goes down or if there are network issues, leading to a complete halt in the training process.

Decentralized Federated Learning: Decentralized federated learning removes the need for a central server, thus mitigating the risk of a single point of failure and potential bottlenecks. In this setup, nodes communicate and share model updates directly with each other in a peer-to-peer fashion. Each node aggregates updates from its neighbors to update its local model. This method promotes robustness and can lead to a more resilient network, as there's no central server that, if failed, would stop the entire process. However, the accuracy and efficiency of the model can heavily depend on the network topology and how well connected the nodes are.

Heterogeneous Federated Learning (HeteroFL): Recognizing the diversity in client devices and data, Heterogeneous Federated Learning, or HeteroFL, accommodates a wide array of devices varying in computational capabilities, storage, and data types, such as mobile phones, computers, and IoT (Internet of Things) devices. Traditional federated learning assumes homogeneity in the devices and data, which is rarely the case in real-world applications. HeteroFL addresses this by allowing these diverse devices to contribute to a global model, acknowledging and utilizing their differences. The goal is to create a robust, inclusive, and adaptable model that performs well across various data distributions and device types. This approach, while complex, provides a pathway to more universally applicable models.

The study by Agleby et al. [219] presents a methodology that employs federated learning (FL) to ensure privacy in training diagnostic models for melanoma, using both skin lesion images and clinical data. They compare the performance of a global federated model with a centralized learning (CL) model, finding that the FL model nearly matches the CL model's F1-Score and accuracy, with slightly higher sensitivity. The study utilized a dataset provided by the Society for Imaging Informatics in Medicine (SIIM) and the International Skin Imaging Collaboration (ISIC). The dataset was divided among five clients, with a combined test set of 412 images, out of a total of 4558 images used for the experiment. The results showed that the Centralized Learning (CL) model had a 5.04% higher specificity in clas-

sifying negatives compared to the Federated Learning (FL) model. However, the FL model outperformed the CL model by 3.27% in sensitivity, correctly classifying more positives. The accuracy and F1 score of the FL model were very close to the CL model, differing by only 0.73% and 0.39%, respectively.

Hossen et al. [220], explore the classification of skin diseases, specifically Acne, Psoriasis, Eczema, and Rosacea, using a dataset of 849 images collected from Atlas-Derm, Derm101, and Dermnet. The methodology centers around employing Convolutional Neural Networks (CNN) with particular emphasis on VGG16 and AlexNet models, integrated with a federated learning framework to enhance privacy and efficiency. The performance of the proposed model indicates an average test accuracy of 83% on federated learning framework. The values for Rosacea are 0% due to the small number of images. While AlexNet and VGG16 did not outperform the proposed model for Acne, Eczema, and Psoriasis, they showed slightly better results for Rosacea.

Lee et al. [221] introduce an Adaptive Personalized Diagnosis Network (APD-Net) was developed to achieve superior skin disease diagnosis. The study utilized three public datasets, 7pt, HAM, and ISIC, alongside a custom dataset comprising 2490 images across Eczema, Dermatitis, Rosacea, and Normal skin. The methodology integrated a Genetic Algorithm (GA)-based fine-tuning method with a dual-pipeline (DP) architecture for deep learning models, enabling personalized and adaptable diagnostics. The GA-based fine-tuning method adaptively customizes the optimized DL model on each edge device, while the DP architecture allows for efficient extraction of feature maps using both generalized and personalized parameters. The APD-Net demonstrated a notable 9.9% improvement in accuracy compared to other models, achieving 88.51% accuracy in federated learning settings.

The study by Wu et al. [222] introduces an on-device FCL framework that facilitates effective learning with limited labels specifically for dermatological disease diagnosis. The FCL framework pre-trains the model on distributed unlabeled data, providing a good initialization, and subsequently fine-tunes with a limited number of labeled data. To enhance data diversity and improve the quality of local contrastive learning without compromising privacy, a feature sharing method is proposed. This method allows diverse features to be shared and contrasted during local learning, leading to better representations. The study presents experimental results showing that the proposed methods outperform state-of-the-art techniques in terms of diagnostic accuracy and label efficiency, using dermatological disease

datasets with various skin colors. The methods were evaluated on four datasets of different skin colors, namely the ISIC 2019 challenge dataset, AtlasDerm, Dermnet, and DarkDerm. These datasets include various types of skin conditions and are used to form a unified classification task across five diseases: basal cell carcinoma (BCC), dermatofibroma (DF), melanoma (MEL), melanocytic nevus (NV), and squamous cell carcinoma (SCC). The total dataset includes about 25k dermoscopic images from ISIC, 11k images from AtlasDerm, 276 images from Dermnet, and 216 images from DarkDerm. The proposed FCL method is evaluated using two settings: local fine-tuning and federated fine-tuning, with various fractions of labels (10%, 20%, 40%, 80%). The mean recall and precision of each class are reported as performance metrics. The pre-training is conducted for 100 communication rounds using FedAvg as the model aggregation algorithm, with the fine-tuning stage involving either 20 epochs in local fine-tuning or 100 rounds in federated fine-tuning. The study claims superior diagnostic accuracy and label efficiency over other state-of-the-art techniques.

Yaqoob et al. [223] propose a privacy-aware machine learning approach for skin cancer detection utilizing asynchronous federated learning and convolutional neural networks (Async-FL-CNN). The method optimizes communication rounds by dividing CNN layers into shallow and deep layers with shallow layers being updated more frequently. A temporally weighted aggregation approach is introduced to enhance accuracy and convergence of the central model. The proposed approach is evaluated on a skin cancer dataset, showing higher accuracy and requiring fewer communication rounds compared to existing methods. The key contributions include improved communication efficiency, leveraging CNNs for clients, and asynchronous federated learning for model aggregation at the global center server. The study used the ISIC-2019 dataset of dermoscopy skin lesion images containing images from eight different classes of skin cancers. The ISIC-2019 dataset contains 25331 dermoscopy images, with 21491 for training, 1930 for testing, and 1910 for validation. The performance of the proposed method was compared with existing baseline federated learning methods in terms of accuracy, loss, precision, communication size, local epoch effects, and convergence rate. The proposed Async-FL-CNN achieves better performance compared with existing and baseline models, as it allows the local clients to asynchronously train their local datasets using CNN. Specific performance metrics such as F1 score, recall, sensitivity, specificity, precision, and loss were used to measure the accuracy of the proposed method, showing that it achieves higher

accuracy rates while requiring fewer communication rounds.

The study by Shi et al. [224] proposes a federated contrastive learning (FCL) framework to enhance the generalizability of diagnostic models for skin lesions in edge computing networks. It introduces a dual encoder network leveraging unlabeled samples for performance improvement and devises a Maximum Mean Discrepancy (MMD)-based supervised contrastive loss function to explore intra-class and inter-class variances of samples. The major contributions include constructing an edge computing-based intelligent skin lesion diagnosis network, an FCL framework with collaborative loss calculation and gradient aggregation, and a new contrastive loss function to efficiently explore complex variances of samples. The proposed FCL framework is verified on the ISIC 2020 dataset, which consists of eight exclusive classes: melanoma (MEL), nevus (NV), seborrheic keratosis, lentigo NOS, lichenoid keratosis, solar lentigo, café-au-lait macule, and atypical melanocytic proliferation. The dataset also includes a substantial amount of unlabeled but benign images. The class distribution is modified to address extreme imbalance, resulting in four classes. The scheme is trained using the ResNet-50 architecture, with the feature vector of the final pooling layer used as the representation vector. The FCL framework was compared with several federated learning schemes such as Federated Averaging (FedAvg), Federated Proximal (FedProx), Momentum Contrast for Unsupervised Visual Representation Learning (MOON), Federated Momentum Contrast (FedMoCo), Federated Semi-Supervised Learning with Inter-Client Consistency and Disjoint Learning (FedMatch), and Federated Exponential Moving Average (FedEMA). The results showed that the presented method exceeded all baselines in different amounts of edge nodes based on the prediction accuracy, forming more representative feature extraction by leveraging massive unlabeled data. The feature sharing mechanism in FCL enlarges the negative sample set in the loss function calculations, thereby strengthening the diagnostic performance

The study by Moon et al. [225] proposes a federated learning approach with a masked attention model to improve the severity classification performance. The model uses masking modeling to overcome data deformation and damage and extracts discriminative severity features from the masked image. The best classification performance was achieved with an F1-score of 0.88 when the masking ratio was set to 0.5. The study used 792 disease images from 44 Korean patients with psoriasis. The images were classified into five severity levels based on the PASI score (Psoriasis Area and Severity Index). Due to class imbalance, the number of images was dif-

ferent for each severity group. The FL-based classification performance confirmed that as the masking ratio increased, classification performance improved up to a certain point. The best performance was observed with an F1-score of 0.88 when the masking ratio was 0.5. Beyond a masking ratio of 0.7, the performance degradation of the model was notable. The results indicate that even with partially lost disease areas, the severity classification performance was satisfactory, with potential for extracting more distinct and differentiated severity features by learning partial disease areas. An ablation study of the masking attention model showed that using fusion attention features made the classification performance more robust.

Furthermore, Adnan et al. [226] proposed differentially private federated learning as a potential method for learning from decentralized medical data, such as histopathology images. Differential privacy enhances this approach by providing quantitative bounds on the privacy level offered. Private federated learning achieves results comparable to conventional centralized training, suggesting its potential for distributed training on clinical data for disease conditions when limited samples available for study. Sheller et al. [227] demonstrated that federated learning across ten institutions yields models that achieve 99% of the quality of those trained with centralized data, also assessing the models' generalizability on data from institutions not part of the federation. Further investigations were made into the impact of data distribution across collaborating institutions on model quality and learning patterns. It was found that the expansion of data access through privacy-conscious, multi-institutional collaborations enhances model quality more significantly than any errors introduced by the collaborative method itself. Additionally, to mitigate bias stemming from local validation set selection, a process termed "collaborative cross validation" was employed. This method ensures a more equitable and accurate validation process across the different participating institutions.

Some key take-aways from the aforementioned studies are:

- While the majority of studies have concentrated on cancerous lesions, Hossen et al. [220] and Lee et al. [221] have extended their research to include clinical images of non-cancerous skin conditions, incorporating Rosacea among others. However, both studies found that the analysis and classification of Rosacea using federated learning were less effective compared to other skin conditions. This underscores the need for proper standardization of Rosacea images to improve their recognition and classification in future federated learning-based studies.

- Wu et al. [222] and Shi et al. [224] have made significant contributions by focusing on Federated Contrastive Learning (FCL), specifically targeting improvements in medical diagnostics and skin disease classification. While both studies aim to advance FCL in medical imaging, Wu et al. focus on the label efficiency and data quality in a distributed setting, whereas Shi et al. emphasize leveraging unlabeled limited data and a unique loss function to improve model generalizability.
- A few studies [219] [226] explored centralized and decentralized architectures to find the most effective strategies for specific scenarios. While some studies emphasize on the centralized approach due to its straightforward coordination and model aggregation process, others highlight the benefits of decentralized Federated Learning.

2.5.5 3D face modelling

For many years, visual computing research has primarily focused on face recognition and representation. The Eigenfaces approach, introduced by Sirovich et al.[228] in 1987 and Turk et al.[229] in 1991, marked significant progress in this field. This method learned face representations from examples and exclusively dealt with grey levels in the image domain. Although the Eigenfaces method was a significant paradigm shift in face recognition, it had some limitations. For instance, it was confined to a particular pose and lighting condition and could not represent shape variations effectively. As a result, when the coefficients in linear combinations of eigenvectors were altered continuously, the model could not find a single parameter for features such as the distance between the eyes, and structures would fade in and out instead of shifting along the image plane.

There have been a few preliminary studies on 3D modelling of faces and other biomedical images using Statistical Shape Models (SSMs) which include Active Shape Models (ASMs) [149] [151] and Active Appearance Models (AAMs) [150]. These have been utilised in preliminary studies for the 3D modeling of faces and biomedical images. Mainly, MRIs of the brain and knees have been considered in depth in these works while other medical applications were also mentioned. The ASM essentially matches a model to boundaries in an image. The AAM finds model parameters that synthesize a complete image similar to the target image by using a set of target feature points. The combination of shape and appearance has turned

out to be very impactful [32]. The better the model represents the structure of the objects to be analyzed, the easier it becomes to fit the model. Due to the linear and parametric nature of SSMs, they are mathematically convenient and easy to work with image analysis algorithms. For this reason, SSMs have become very popular [152]. However, SSMs come with various drawbacks such as (a) the shape variation in SSMs is restricted to the linear span of the training data, (b) hence, a lot of training data is needed.

The application of computer vision for face modeling has focused on the direction of a new face representation using analysis-by-synthesis [230]. This seeks to explain an image by synthesizing its content using both 2D and 3D modelling. One technique which does this is 3D Morphable Models (3DMMs)[31].

A 3D Morphable face model is based on the two fundamental ideas (a) all faces are in dense point-to-point correspondence (which is usually established on a set of example faces in a registration procedure), (b) this correspondence can be maintained throughout any further processing steps[32]. A 3DMM is a generative model which applies to the entire shape of the face and the appearance of the face. The 3DMMs were derived from textured 3D scans of 100 females and 100 males. Subsequently, fitting was carried out using 6,428 publicly accessible facial datasets (from the CMU-PIE and FERET database), with an emphasis on capturing multi-pose, illumination, expressions, and full-frontal face views. This study facilitated modeling of various ages, ethnic groups, and facial expressions.

2D morphable models and 3D morphable models rely on dense correspondence rather than only a set of facial feature points[32]. Determining the dense correspondence is only possible by assigning every point on the reference object that is semantically meaningful to the corresponding point on the target object. This process is called image registration. The same anatomy of an object can be explained using any other object of the same class perturbed with deformation or slight structural variation[231].

To encourage research in parametric models that are generative in nature, in 2009, Paysan et al.[232] introduced Basel Face Models (BFMs), which offers a higher shape and texture accuracy than the previous 3DMMs due to a better scanning device and fewer correspondence artefacts. They are a variant of 3DMMs and are publicly available.

Despite the significant advancements in 3DMMs, there is still room for research on extending them to models that do not have regular faces. Applying 3DMMs to



Figure 2.3: Standard reference shape. This image is taken from the work by Lüthi et al.[152]

issues requiring a large aggregate of details and complexity often results in a less efficient representation. This makes it challenging to determine the precise number of parameters necessary to represent both the facial appearance and geometry at different levels of detail in facial aesthetics. Although 3DMMs have shown promising results in various applications, the limitations in fitting mean that they do not offer enough freedom for designing deformations beyond the regular facial structures/aesthetics. Overcoming this problem requires a non-rigid approach to image registration.

The generalization of SSMs is called Gaussian Process Morphable Models (GPMMs)[152]. The application is based on the Principal Component Analysis (PCA) concept with a covariance function computed from the training data. The GPMMs start with standard reference shapes such as shown in Fig. 2.3.

Later, in order to incorporate a specific family of deformations with a dataset, a framework is needed which can assign the probability to all possible deformations for the given input feature. The desired deformations can be modelled by defining a covariance function as shown in Fig. 2.4.

According to Luthi *et al.* [152] GPMMs come with many advantages over SSMs such as:

- With GPMMs, there is much more freedom in defining the covariance function, by combining different covariance functions (or kernels) to mimic more sophisticated registration schemes. This helps to extend the model beyond the linear span of example data (training data), hence GPMMs work well with little training data.
- GPMMs are generative; therefore, the validity of prior assumptions can be



(a) $s = 100, \sigma = 100$ mm



(b) $s = 10, \sigma = 30$ mm

Figure 2.4: Generated face models using a Gaussian Covariance function with different standard deviations. This image is taken from the work by Lüthi et al.[152]. The model parameters include the scale factor s and the bandwidth σ , where s determines the variance (i.e., scale) of a deformation vector. Choosing a large σ leads to smooth, global deformations of the face, whereas a smaller σ yields more localized deformations.

assessed by sampling from the model.

- Shape variations can be well approximated using only a moderate number of leading basis functions (eigenvectors).
- For most anatomical shapes, finely detailed deformations only occur in parts of the shape, and GPMMs give more power to model these slight deformations only where they are needed.
- The above advantages may help in incorporating the medical expert knowledge into the model in order to shape slight deformations.

Hence, GPMMs can be incorporated to deal with the significant and minor deformations that occur on human faces along with various subtypes of rosacea (i.e. phymatous rosacea or *Rhinophyma*) by creating a 3D model of faces with possible deformations from the set of 2D images. GPMMs may give the modelling power to model these fine deformations only where they are needed. This autonomy may

help in incorporating medical expert knowledge in order to model slight, local deformations e.g. lesions, enlargement of nose etc. As the effective treatments of rosacea are broadly advancing, a laser and light-based treatment approach to the rosacea diagnosis is recommended[233], [234]. Thus, reconstructing local deformations on the face for specific subtype-3 of rosacea may help diagnose and support light and laser treatment of the disease condition.

As GPMMs are generative, recently there have been a few applications developed using CNNs and GANs. The image registration experiments are generally conducted using Statismo [235, 152]. Scalismo [152], on the other hand, was employed for tasks such as model construction, surface registration, and Active Shape Model fitting. Primarily, Scalismo is devised to generate and analyze statistical models concerning shape and appearance, finding applications in fields like medical imaging and computer vision.

Parametric models have also succeeded in generating 3D faces [154], [236], [153], [237], [45], [238], [156], [239]. Therefore, if local deformations of rosacea are successfully generated using Parametric Modelling, then it may help in generating more synthetic datasets of rosacea using GAN models.

2.6 Discussion and outlook

The majority of the research done in the field of skin disease diagnosis is focused on skin cancer. The lack of publicly available visual data for a skin condition such as rosacea often leads to poor performance in classification and automated diagnosis models. However, the recent advancement in the field of AI and particularly data generation has prompted many opportunities in the future of computer-aided diagnosis for Teledermatology[240] including rosacea.

Although data augmentation and transfer learning have been very successful with medical and clinical image analysis with large datasets, they may not perform as accurately with limited data. Hence, using a classification approach for a limited data problem may not be a good idea at present. However, there is an opportunity to explore techniques to overcome this central problem which may improve the scope of research.

In the field of skin disease analysis, generating synthetic samples that may look as real as certain skin conditions, such as rosacea, can mitigate the problem of data scarcity. As discussed in Section 2.5.2., a few variants of StyleGAN can be utilised

to explore the possibility of generating synthetic faces with rosacea. Nevertheless, the accuracy of the generated data must be examined. For such examinations, it is important to rely on the subjective evaluation methods for image quality, such as the Mean Opinion Score (MOS) of a group of experts.

As discussed in Section 2.5.3, GANs have a few common limitations, which may lead to unsatisfactory outcomes. Therefore, it is essential to investigate additional methods to deal with limited data without modification, i.e. by keeping the data volume constant and still performing classification. This can be achieved with hyperparameter optimization by adapting the Meta-Learning concept. Nevertheless, meta-learning, few-shot classification and federated learning are new approaches in medical and clinical image analysis that offer a solid motivation to explore the data scarcity problem. There are a few studies on meta-learning and federated learning for clinical skin disease image analysis, which may provide limited scope for exploration. Further, we discussed 3D modelling as an approach to generate synthetic facial data through GPMMs. By leveraging the concept of GPMMs, unlike GANs, we can handcraft the particular types of appearances on the skin caused by rosacea. This may help identify a few subtypes of rosacea that cause significant deformation on the face.

2.7 Conclusion and Future scope

This chapter has provided a comprehensive overview of the current state of AI and computer vision in the context of skin disease diagnosis, specifically focusing on the challenges and opportunities in handling skin disease image data. The discussion particularly centered on addressing the limited data challenges associated with Rosacea skin condition. We have delved into the effectiveness of deep learning and its remarkable achievements in health data analysis, while also highlighting the critical gaps and challenges that remain, particularly in the context of limited data availability.

The exploration of advanced techniques such as Data Augmentation, Transfer Learning, GANs, Meta-Learning, and Few-shot classification, Federated Learning as well as 3D face modelling using GPMMs, has laid a foundation for understanding the varied approaches available to leverage limited data in the field of dermatology. These techniques have been discussed not only for their potential but also in light of the significant research gaps and noteworthy advantages they present, offering a

balanced view of the current landscape.

As this review serves as a background and literature review for the thesis, it sets the stage for the subsequent research and analysis presented in later chapters. It contextualizes the importance of each technique within the broader aim of improving computer-aided diagnosis for skin conditions and establishes a clear direction for future research endeavors. The recognition of the fundamental challenge of drawing valid inferences from small datasets in machine learning and computer vision underscores the importance of continued innovation and exploration in these areas.

Moving forward, Chapter 3 and 4 will focus on further exploring and refining methodologies of synthetic data generation for skin disease analysis, particularly for conditions like rosacea subtypes. This will include a detailed assessment of the techniques such as GANs and 3D face modelling and their potential to enhance the quality and effectiveness of AI models. By advancing these synthetic data generation technique, we aim to contribute to the development of more robust, accurate, and accessible data and diagnostic tools, ultimately improving patient care and outcomes in dermatology.

Chapter 3

High-Fidelity Synthetic Rosacea Data Generation

3.1 Introduction

As explored in the earlier chapters of this thesis, the diagnosis and understanding of skin diseases through computer vision and machine learning techniques have taken a significant leap forward with the advent of computer-aided diagnosis systems. This chapter builds upon the foundation laid in the previous discussions, focusing specifically on the challenges and innovations in the field of Generative Adversarial Networks (GANs) for skin disease image generation, with a particular emphasis on *Rosacea*.

The introduction of deep learning architectures like Inception v3 [95] has revolutionized the accuracy of skin cancer classification, utilizing large datasets to achieve remarkable precision [82, 241]. However, as noted in the literature review, the limited availability of extensive datasets is a significant barrier, especially for non-fatal chronic skin conditions such as *Rosacea*. While the focus and data accumulation for skin cancer are extensive due to its severity, there exists a stark contrast in the data availability for conditions like *Rosacea*, which, despite their impact on patients' quality of life, have not seen comparable data collection efforts.

The generation of synthetic data by deep generative algorithms, mirroring the characteristics of authentic data is an innovative approach to circumvent data scarcity [242]. This chapter aims to address these disparities by investigating the potential of GANs to augment the limited datasets available for such conditions. Specifically,

it will delve into the use of synthetic data generation to expand the small existing datasets for Rosacea, employing a variant of the StyleGAN architecture [190, 194] trained on only 300 images. By generating high-quality, diverse synthetic images, this approach seeks to enhance the training and subsequent performance of Deep Convolutional Neural Networks (DCNNs) in the future, even when real-world data is scarce.

Moreover, while prior studies [180, 186, 187, 44, 188, 189] have successfully generated synthetic images of skin cancer lesions using various GAN architectures, the focus has predominantly been on localised regions captured through dermatoscopes or similar devices. This research contrasts with previous studies by utilising full-face images to comprehensively capture the manifestation of *Rosacea*. This distinction is crucial, as *Rosacea* affects various facial regions and understanding its subtype-specific impact requires a broader, more inclusive imaging approach. In this chapter, the focus will be on generating and analysing synthetic images for Subtypes 1 (*Erythematotelangiectatic Rosacea*) and 2 (*Papulopustular Rosacea*), noting the progressive nature and potential transitions between these subtypes [14].

The subsequent sections of this chapter will detail the methodologies employed, the challenges encountered, the results obtained, quantitative and qualitative evaluations of the generated images and the implications of these findings for future research and clinical practice. By the end of this chapter, the reader will have a comprehensive understanding of the role of GANs in generating high-fidelity synthetic images, particularly for *Rosacea*, setting the stage for the continued advancement of synthetic data generation, effective validation and computer-aided diagnosis systems in dermatology.

3.1.1 Importance of studying full-face images of *Rosacea*

Generally, prolonged redness is one of the common early symptoms (*pre-Rosacea*) that usually appears over the cheeks, chin, nose, or forehead. Eventually, certain patients develop some swelling (*'edema'* in medical terminology), which is noticeable at the very early stage of the disease. Particularly, the locality and the visible regions of blood vessels (*'telangiectasia'* in medical terminology), prolonged redness and edema may give the impression of the severity of the disease.

Full-face images of patients have proved very useful in diagnosing the condition and predicting the timeline of growth for future diagnosis and treatment. Hence, in this research, we are considering full-face images of *Rosacea*, while most of the

previous studies mentioned in Chapter 2 have considered only partial images of the face.

To summarize, our **contributions** are as follows:

1. In this chapter, to the best of our knowledge, a small dataset consisting of 300 full-face *Rosacea* images is utilized for the first time as a basis for generating synthetic images.
2. We show how fine-tuning the model (StyleGAN2-ADA) and varying experimental settings significantly affect the fidelity of *Rosacea* features.
3. We demonstrate that R_1 Regularization strength ' γ ' helps achieve high-fidelity characteristics of *Rosacea* condition.
4. We generate 300 high-fidelity synthetic full-face images with *Rosacea*, which can be further utilized to expand the *Rosacea* face dataset for computer-aided clinical diagnosis.
5. We present qualitative evaluations of synthetic/generated faces by expert dermatologists and non-specialist participants. These show the realistic characteristics of *Rosacea* in generated images.
6. We critically analyse the quantitative evaluation such as validation metrics(s) from the list of conducted experiments and point out the limitations of usage of validation metric(s) alone as evaluation criteria in the computer-aided medical image diagnosis field.

3.2 Developments in Synthetic Face Generation

Followed by GANs, the introduction of Deep Convolutional GANs (*DCGANs*) marked significant advancements in the field, setting a foundation for the evolution of more sophisticated models in synthetic face generation. The limitations of DCGANs, such as model instability, mode collapse, filter leakage after prolonged training time, and the generation of images at smaller resolutions, highlighted essential areas for further research and development in GANs. These limitations and the quest for higher quality and resolution in generated synthetic faces paved the way for the development of Progressive Generative Adversarial Networks (Progressive GANs).



Figure 3.1: Progress of synthetic face generation using various GAN models with the maximum volume of dataset available.

The Progressive Growing of GANs (*ProGANs*) introduced by Karras et al.[183], improved the resolution of the generated images with a stable and swifter training process. The main idea of ProGANs is to start from a low resolution i.e. 4×4 and then progressively increase the resolution, e.g. up to 1024×1024 , by adding layers to the networks. The training time is 2-6 times faster depending on the desired output resolution. ProGANs can generate 1024×1024 facial images using the CelebA-HQ[183] dataset with 30,000 selected real images in total. The idea of ProGAN emerged from one of the GANs architectures introduced by Wang et al.[191]. Although ProGAN successfully generated facial images with large resolution, it did not work adequately in generating realistic features and microstructures.

Although the generation of high-resolution images was achieved by GANs, there were still indispensable research gaps that needed to be addressed. Thus, the introduction of *StyleGAN*[190] came with further improvements which helped in understanding various characteristics and phases in synthetic image generation/image synthesis. Important improvements in the StyleGAN architecture include:

- Upgrading the number of trainable parameters in style-based generators; this is now 26.2 million, compared to 23.1 million parameters in the ProGAN[183] architecture.
- Upgrading the baseline using upsampling and downsampling operations, increasing training time and tuning the hyperparameters.
- Adding a mapping network and adaptive instance normalization (AdaIN) operations.
- Removing the traditional input layer and starting from a learned constant tensor which is $4 \times 4 \times 512$.
- Adding explicit uncorrelated Gaussian noise inputs, which improves the generator by generating stochastic details.

- Mixing regularization which helps in decorrelating the neighbouring styles and taking control of fine-grained details in the synthetic images.

In addition to the improvements in generating high-fidelity images, StyleGAN introduced a new dataset of human faces called Flickr Faces HQ (FFHQ). FFHQ has 70,000 images at 1024×1024 resolution and is of a diverse range of ethnicity, age, background artifacts, make-up, lighting, image viewpoint and various accessories such as eyeglasses, hats, sunglasses etc. Based on these improvements, comparative outcomes are evaluated using a metric called Fréchet Inception Distance (FID)[243] on two datasets i.e. CelebA-HQ[183] and FFHQ. The recommended future investigations include separating high-level attributes and stochastic effects while achieving linearity of the intermediate latent space.

Successively, another variant of StyleGAN was introduced by Karras et al. called *StyleGAN2*[43], in which the key focus was exclusively on the analysis of the latent space denoted as W . The latent space W represents a more disentangled and intermediate latent space that allows for more controlled and interpretable variations of generated images, compared to the direct noise input space typically used in GANs. This refined latent space enables StyleGAN2 to produce higher quality and more diverse synthetic images by manipulating attributes more precisely. As the generated output images from StyleGAN contained some unnecessary and common blob-like artifacts, StyleGAN2 addressed the causes of these artifacts and eliminated them by defining some changes in the generator network architecture and in the training methods. Hence the generator normalization is redesigned, and the generator regularization is redefined to boost conditioning and to improve output image quality. The notable improvements in the StyleGAN2 architecture include:

- The presence of blob-like artifacts such as those in Fig. 3.2 was addressed by removing the normalization step from the generator in the network architecture. This modification to the generator’s structure mitigates the formation of such artifacts, contributing to a cleaner and more accurate image generation. The removal of normalization helps in stabilizing the feature scale across different layers, effectively reducing the likelihood of these unwanted visual anomalies. (Generator redesign).
- Grouped convolutions are employed as a part of Weight demodulation, in which weights and activation functions are temporarily reshaped. In this setting, one convolution sees one sample with N groups, instead of N samples

with one group.

- Adaption of *Lazy Regularization* in which R_1 regularization is performed only once in 16 mini-batches. This reduces computational costs and memory usage in total.
- Adding a path length regularization aids in model reliability and performance. This offers a wide scope for exploring the architecture at the further stage. Path length regularization helps in creating denser distributions without mode collapse problems.
- Revisiting the ProGAN architecture to adapt benefits and remove the drawbacks e.g. progressive growing in the residual block of the discriminator network.

The datasets LSUN[244] and FFHQ were used with StyleGAN2 to obtain quantitative results through metrics such as FID[243], Perceptual Path Length (PPL)[190], Precision and Recall[245].



Figure 3.2: An example of blob-like artifacts in the generated images. This image is taken from Karras et al.[43]

Another set of GAN architectures called *BigGAN* and *BigGAN-deep*[246] expanded the variety and fidelity of the generated images. These improvements included making architectural changes which improved scalability, a regularization scheme to recuperate conditioning as well as to boost performance. The above modifications gave a lot of freedom to apply the “*truncation trick*”, a sampling method that aids in controlling sample variety and fidelity in the image generation stage. Even though different GAN architectures produced improved results over a period, model instability during training was a common problem in large-scale GAN architectures[247]. This problem was investigated and analysed through the introduction of BigGAN by leveraging the existing techniques and by presenting novel techniques. The ImageNet ILSVRC 2012 dataset[141] with the resolutions 128×128 , 256×256 , 512×512 was used in BigGAN and BigGAN-deep architectures

for demonstrating quantitative results through metrics such as FID and Inception Score (IS)[248].

The aforementioned GAN architectures were trained on a large amount of data and can generate high-resolution outputs with variety and a fine-grained texture. Therefore, it is important to expand the potential of GAN architectures to perform well and produce high-fidelity synthetic images, even if there are limited images available.

However, the key problem with a small number of images is the overfitting of training examples in the discriminator network. Hence the training process starts to diverge, and the generator does not generate anything meaningful because of overfitting. The most common strategy to tackle overfitting in deep learning models is “*data augmentation*”. There are instances in which augmentation functions learn to generate the augmented distribution, which results in “*leaking augmentations*” in the generated samples. The leaking augmentations are the features which are learned from the augmentation style rather than the features which are originally present in the real dataset.

Hence to prevent the discriminator from overfitting when there is only limited data available, a variant of StyleGAN2 called *StyleGAN2-ADA*[194] has been introduced with a wide range of augmentations. An adaptive control scheme was presented in order to prevent such augmentations from leaking in the generated images. This work produced promising results in generating high resolution synthetic images obtained with a few thousand images. The significant improvements in StyleGAN2-ADA include:

- *Stochastic Discriminator Augmentation* is a flexible method for augmentation that prevents the discriminator from becoming overly confident by showing all the applied augmentation to the discriminator. This assists in generating desired outcomes.
- Addition of *Adaptive Discriminator Augmentation (ADA)* by which the strength of augmentation ‘ p ’ can be adjusted at every interval of 4 mini-batches N . This technique helps in achieving convergence during training without any occurrence of overfitting irrespective of the volume of the input dataset.
- *Invertible transformations* are applied to leverage the full benefit of the augmentation. The proposed augmentation pipeline contains 18 transformations grouped in 6 categories viz. pixel blitting, more general geometric transfor-

mations, colour transforms, image-space filtering, additive noise, and cutout.

- Capability to handle small-volume datasets such as, 1000 and 2000 images from FFHQ dataset, 1336 images of METFACES [249], 1994 overlapping cropped images from 162 breast cancer histopathology images called BRECAHAD[250], nearly 5,000 images of AFHQ and 50,000 images of CIFAR-10 [251].
- Although the small volume of the dataset is the main feature in the StyleGAN2-ADA, some high-volume datasets are broken down into different sizes for monitoring the model performance. The FFHQ dataset is used for training the model. Various subsets of the dataset such as 140,000, 70,000, 30,000, 10,000, 5000, 2000 and 1000 are used to test the performance. Similarly, dataset LSUN CAT is considered with the volume starting from 200k to 1k for model evaluation. The FID is used as an evaluation metric for comparative analysis and demonstration of StyleGAN2-ADA model performance.

Amongst the studies and related work regarding face generation using GANs as discussed above and represented in Fig. 3.1, StyleGAN2-ADA appeared to work adequately with a small volume of data. Particularly in the case of small volumes of medical/clinical images, StyleGAN2-ADA is a useful method for investigation. Considering the advantages of StyleGAN2-ADA, in this research, we implemented and trained the model with 300 images of *Rosacea*, to be discussed in section 3.4.

3.3 Methodology

3.3.1 StyleGAN2 with Adaptive Discriminator Augmentation

The above analysis of the state-of-the-art techniques indicates that StyleGAN2-ADA can potentially be used to address the data scarcity of *Rosacea* by generating synthetic samples.

The most attractive point of StyleGAN2-ADA is its ability to handle a small amount of data, in fact a minimum of 1000 images. This is achieved by utilizing the concept of *Adaptive Discriminator Augmentation (ADA)*.

The concept of *ADA* is motivated by three well-known **limitations** of GAN models [252][253]:

1. Difficulty in handling small amounts of data.
2. Discriminator overfitting which leads to mode collapse.
3. Sensitivity to the selection of hyperparameters.

Generally, when condition 1 exists, it is more probable for condition 2 to occur, and when both exist, it leads to catastrophic failure in most GAN models. Nevertheless, when limited data is available, one possible solution for overfitting is “Data Augmentation”. Data Augmentation helps in expanding the input images by applying temporary alterations such as geometric transformations and preprocessing tasks. This practice helps in increasing input feature space during the training.

However, the employment of augmentations can pose challenges. Many of the existing GANs models augment the real images during training, leading the discriminator to learn these augmentations as part of the real image distribution. The crux of the challenge is ensuring the generator accurately discerns the inherent distribution of the actual data and does not assimilate the specific distortions introduced by the augmentations. If the discriminator never observes the true distribution of the training images, its capability to adequately guide the generator becomes questionable. The aim is to navigate the generator towards comprehending the intrinsic patterns and variations inherent in the real data, avoiding the characteristics imprinted by the augmentations [254]. Hence the generator learns to produce images with undesired augmentation artifacts such as noise, colour, cutout, and geometric operations. This learning practice and producing images with undesired augmentation artifacts are called “*leaky augmentations*”.

A wide range of augmentations may be used to stop the discriminator from overfitting while ensuring that applied augmentations do not leak into the resulting generated images. In addition, an Adaptive control procedure may enable the model to function effectively irrespective of the volume of training data, the dataset’s nature/characteristics, and the training approach.

Overfitting in various GAN models, especially in the variants of StyleGANs, can be observed when the value of the Fréchet Inception Distance (FID)[243] metric starts to escalate without any decline, leading to leakage in the augmentations. The phrase ‘escalate without any decline’ refers to the observed trend of the FID metric continuously increasing, indicating a deterioration in the quality of generated images. Typically, a stable or decreasing FID score suggests improving or stable model performance. However, in the context of overfitting, the FID begins to rise, reflect-

ing that the generated images are diverging from the desired distribution of real images. This ‘escalation’ of the FID metric, particularly without any subsequent decrease, signals that the model is beginning to memorize the training data rather than learning to generalize, leading to poorer performance on unseen data. To prevent such behaviour, a pipeline known as “*Stochastic Discriminator Augmentation*” is introduced. This approach is inspired by the balanced consistency regularization (bCR) approach by Zhao et al.[255], designed to prevent leaking of the augmentations. Stochastic Discriminator Augmentation is a flexible type of augmentation that prevents the discriminator from becoming overly confident by showing all the applied augmentation to the discriminator. The discriminator is evaluated based on the augmented images, using the same augmentation as was applied when training the generator. In this practice, the discriminator can see the training images, which assists the generator in generating the desired outcome. Fig. 3.3 shows the workflow of Stochastic Discriminator Augmentation.

Similarly, in order to regulate the distribution in the generated images, the idea of *invertible transformation* is used. Invertible transformations are beneficial when applying a wide range of augmentations; for example, 18 types (clustered into 6 categories) of augmentations are used. Invertible transformation in the augmentation can be defined as, “for a target distribution y and an augmentation operator T , the generated distribution x is trained such that the augmented distributions match with the target distribution y ” [194]. If a transformation is non-invertible there will be leakage but if all the transformations are invertible there will be no leakage. Invertible transformations can be reversed by the generator and removed from the distribution while non-invertible ones cannot be removed and can result in leakage. The generator network learns to generate the images in the correct underlying distribution by undoing the augmentation that does not fit the right kind of distribution. Hence, applying this concept of invertible transformation in augmentation [256] helps with finding the correct target distribution of the data.

Another trick used to prevent leaking is to apply different augmentations in *a particular fixed order*; for example, blitting, geometry and colour. Therefore, a sequential composition of augmentations that do not leak will ensure no leakage to the generated images.

Although the Invertible Transformation process prevents the augmentation from leaking at least at the very early stages of training, which is desirable, a few constraints still require to be addressed. Augmentation leaking is highly dependent on

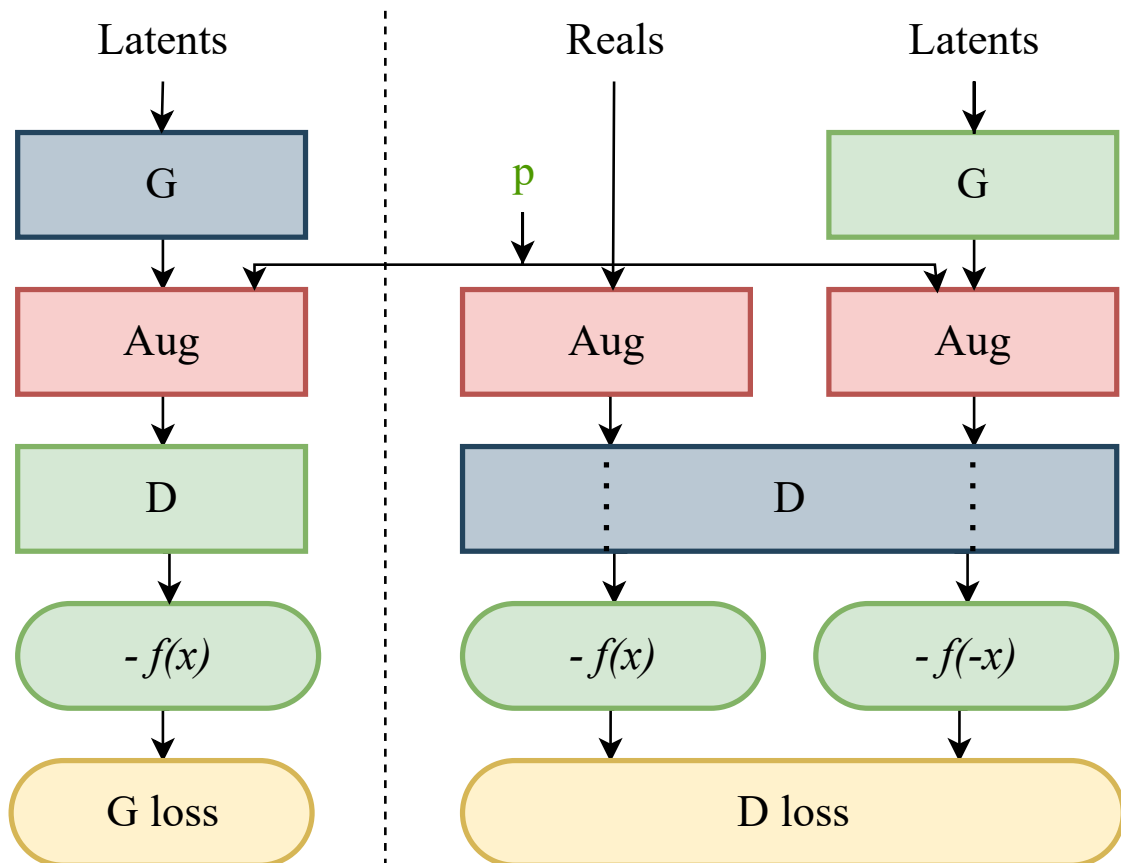


Figure 3.3: The flow diagram of Stochastic Discriminator Augmentation [194], where G is the generator and D is the discriminator. The red boxes represent the 18 augmentation operations. The set of selected augmentation are controlled by the augmentation probability ' p ' and these augmentation can be visible to the discriminator D in the green box. The blue boxes represent the networks that get trained during the training process and yellow boxes represent the loss calculated after the training. In this set up, the non-saturating logistic loss is accommodated to calculate the final probability of the images being predicted as fake.

a probability value, ‘ p ’. Higher values of ‘ p ’ may confuse the generator by picking one of the random possibilities of the augmentation and image distribution; this phenomenon makes the chosen augmentations leak. If ‘ p ’ is under a safety limit, it is less likely to produce leaking augmentation on generated images. To keep track of the safety limit value, an adaptive approach was introduced.

The concept of *Adaptive Discriminator Augmentation* was supported by *controlling the augmentation strength ‘ p ’* by which the augmentation is applied as the training progresses. The initial value of ‘ p ’ starts from 0 and gets regulated in every 4 mini-batches as training progresses. If overfitting occurs during the training, the p-value can be adjusted by a fixed rate. A given target value can control the strength of the p-value. This concept of setting a target value, aka “*ADA target*”, came from observing the training process and the safety limit of value ‘ p ’. For example, In the study by Karras et al. [194] it was observed that the FID value declined after ‘ p ’ became close to 0.5. Hence the ADA target was set as 0.6. Regardless of the dataset volume, discriminator overfitting was avoided by implementing this strategy, and convergence was achieved during the training.

Despite the fact that GAN models are very sensitive to hyperparameter selection, StyleGAN2-ADA supports reasonable quality of results without major changes in the hyperparameters and loss functions while training from scratch or performing transfer learning.

3.3.2 The impact of R_1 Regularization γ for 300 images

As discussed in section 3.3.1., one of the limitations of GANs is that small data may lead to overfitting, divergence or mode collapse. These grounds motivate our work to adapt StyleGAN2-ADA, which uses a minimum of 1000 images for experimental purposes. In this work, we used a limited amount of input images i.e. 300 images, but with fine-grained vital features i.e. *Rosacea* condition. Given the limited number of images, it might be hard to retain the most important features while training the networks and generating synthetic images. Hence, it is necessary to explore the strategies which may help obtain better results along with the adaptation of StyleGAN2-ADA.

The StyleGAN2-ADA architecture functions very well even without changing network architectures, loss functions or other key parameters. As GANs are sensitive to hyperparameters, in this work most of the hyperparameters are kept unchanged except for the R_1 Regularization weight/strength ‘ γ ’. According to a few studies,

regularization has a significant impact on stabilizing GAN training. Regularization helps produce high-quality images by stabilizing the broad range of noise levels [257]. During the training of GANs, the generator takes random noise as input and produces synthetic images. However, without proper control, this random noise can lead to high variance in the generated images, manifesting as unrealistic or highly varied outputs. Regularization techniques like R_1 Regularization impose constraints on the model's parameters, encouraging smoother, more stable learning. In the case of StyleGAN2-ADA, R_1 Regularization helps in managing the influence of noise by penalizing large, erratic shifts in the model weights, leading to more consistent and controlled image generation. It effectively reduces the model's sensitivity to the specific noise patterns in the input, resulting in higher quality images that are less prone to artifacts or distortions commonly associated with unstable GAN training.

In the instances of images with a high number of features, R_1 Regularization (aka L_1 norm Regularization) performs satisfactorily in feature selection by removing some unimportant features. It helps in shrinking the coefficient of the less important features to 0. R_1 Regularization helps to prevent overfitting. To prevent overfitting due to the small volume of data, regularization extensively reduces the variance of the model without losing important attributes in the input image features and without a significant rise of bias in the model. On the contrary, if the strength value ' γ ' is set too high, the model can miss important details from the input images. In this work, those particular numerical values of ' γ ' are explored, with the aim of retaining vital details of the input images.

As the GANs concept is based on the *Zero-sum game*, it is expected to attain a *Nash Equilibrium* in which each player cannot reduce their cost function without changing the parameters of the other player[258]. As defined, equilibrium is a situation in which no player could improve its position by choosing an alternative available strategy('cost function' in this case), without implying that each player's privately held best choice will lead to a collectively optimal result [259].

The *cost/loss/value function* is affected by the integrated R_1 Regularization. It is necessary to achieve the lowest divergence between the training distribution and the model distribution that obtains minimum loss at equilibrium. Despite this, it is hard to reach the closest point towards the equilibrium when the input images are short in supply. Hence it is essential to leverage the advantage of R_1 regularization strength to achieve minimum loss. The study by Mescheder et al.[260] stated that using R_1 regularization helps in stable training as well as high-resolution image

distribution for CelebA and LSUN datasets. Under suitable assumptions the R_1 regularization strength has an impact on obtaining notably better results in the generated image quality.

L_1 regularization (or R_1 Regularization) is added to the cost function of GANs (in Equation 2.1) as:

$$R_1 = \gamma \sum_{i=1}^n |w_i| \quad (3.1)$$

Where,

- γ is the regularization strength that decides the amount of regularization to be applied.
- $|w_i|$ is the absolute value of each weight in the model, which forces the smaller weights towards zero and hence reduces model complexity.
- n represents the number of parameters in the model.

When the R_1 regularization term is integrated into the original cost function, the new cost function becomes:

$$J_{\text{reg}} = J + R_1 \quad (3.2)$$

To minimize this, we need to consider both terms. When taking the derivative with respect to the parameters (for Gradient Descent), it will involve the derivatives of both the original cost function and the regularization term. The regularization term's derivative will have a component from the sign of the weight, enforcing the sparsity.

Using gradient descent methodologies, each parameter w_i is iteratively adjusted using:

$$w_i(t+1) = w_i(t) - \alpha \left(\frac{\partial J}{\partial w_i} + \gamma \cdot \text{sign}(w_i) \right) \quad (3.3)$$

Where:

- α is the learning rate determining the step size in the direction opposite to the gradient.
- $\frac{\partial J}{\partial w_i}$ is the partial derivative of the original cost function with respect to the weight w_i .

The optimal regularization strength γ and learning rate α need to be tuned carefully via techniques like cross-validation to achieve a trade-off between model fit and model complexity, ensuring generalized and stable models. This nuanced interplay of mathematical operations and strategic integration of regularization in the cost function helps in achieving balanced, robust, and efficient GAN models. It is this intricate math that empowers the model to learn and generalize effectively, producing high-quality synthetic images that are nearly indistinguishable from real ones.

Hence this work examines the effects of R_1 regularization to find the most favourable strength ' γ ' that suits the nature of our dataset, since choosing the value of ' γ ' is highly dependent on the dataset size and its nature. However, a few studies have proposed a mathematical formulation to initiate the value of ' γ ' as an initial guess in Equation 3.4; Where $N = w \times h$ (in this case 512×512) and M is the size of minibatch, w and h are the number of pixels[194, 257]. In the study by Mescheder et al.[261] on the impact of regularization, even though only a handful of images were used, the authors proved that an appropriate choice of γ leads to better convergence properties near local Nash-equilibrium, which further leads to the generation of high-fidelity images while preserving fine-grained details learned from the input images.

$$\gamma_0 = 0.0002 \cdot N/M \tag{3.4}$$

3.3.3 *Rosacea* Data

GANs have produced impressive results due to the availability of the enormous volume of images on various web sources, which have relaxed terms of privacy and copyright. Most of the large datasets used in the improvement and study of GANs contain objects, animals, paintings or faces of celebrities. StyleGAN2-ADA uses histopathological images of breast cancer, which do not disclose patients' identities. Similarly, some other imaging modalities such as dermoscopic imaging, X-Ray imaging, and MRI scans may not disclose the person's identity. Especially when a skin condition is captured directly, focusing on the affected region of the body, a person's identity is hardly identifiable. However, in the case of full facial images with skin conditions such as *Rosacea*, capturing the entire face can result in identifying the patient.

Publicly Available Data

A few teledermatology web sources support computer-aided skin disease diagnosis research and development. The available *Rosacea* images in various web sources are listed in the Chapter 2 of this thesis (same has been mentioned in Mohanty et al.[262]). There are about 208 *Rosacea* images in total. Among these there are only a few images with full-face visibility and a few others are watermarked, which may affect the features in the generated images. In order to examine the nature of *Rosacea* in the facial region, it is essential to access high-quality full face images which are rarely found in online teledermatology sources. Hence, acquiring full-face images of *Rosacea* is a difficult task.

***Rosacea* Dataset- ‘rff-300’**

In this chapter, we have access to a small dataset, which is referred to as the “Irish Dataset” in the rest of the study. The “*Irish Dataset*” is provided by *The Powell Lab, Charles Institute of Dermatology, University College Dublin*[263, 264]. The dataset contains 70 high-quality full-face images of *Rosacea*. The original images were present in various resolutions ranging from 800×1000 to 900×1200 . These were later resized for the experiments. Among the 70 images in the Irish Dataset, 67 images were selected for experiments conducted in this research.

Given the low number of images in the Irish dataset, it was essential to collect more data from various web sources, i.e. the teledermatology web sources and other Google search results. Thus, another 67 full-face images have been taken from *SD-260*[126]. A few more images were obtained from Google search results and teledermatology websites in accordance with the following criteria/standards:

- The resolution is a minimum of 250×250 .
- visibility of full face including forehead to chin and both cheeks.
- The images are labelled/captioned/described under subtypes 1 and 2.

Given these standards/criteria, many *Rosacea* labelled images with partially visible faces are not considered in this research study. This data gathering results in a total number of 300 real-world images for the experiments to generate synthetic full-face images with *Rosacea*. These 300 images are full front-view facial images with *Rosacea* subtype 1 or subtype 2.

All 300 images are centre-cropped manually while preserving the visibility of the face and eliminating unnecessary background details and accessories around the ears and heads. The images are resized to 512×512 pixels to keep the optimum details of the disease. The preferred file format type “.png” was chosen to preserve the best possible sharpness of the original images. For ease of understanding and usage, the entire dataset used in the experiments is referred to as “**rff-300 (Rosacea-full-face-300)**”.

Implementation Specifications

A system equipped with an Nvidia Geforce RTX 3090 (24GB) GPU, an AMD Ryzen 9 5900X 12 core CPU, and 32 GB RAM are used to carry out the experiments. The complete implementation was carried out on Pytorch 1.7.1. with CUDA version 11.1 on Linux.

3.4 Experiments and Results

The implementation choices in this work are the same as in the original work on StyleGAN2-ADA with some minor changes in the configuration. As the original work claims to have chosen the ideal configuration in network architecture and loss functions, these units are kept unaltered in these experimental implementations. The learning rate of 0.0025 is kept unchanged to examine the effect of augmentation and other existing hyperparameters on the output. All the 300 input images with resolution 512×512 are x-flipped, which brings the number of input images to 600.

In most cases, the augmentation choices are limited to pixel-blitting and geometric augmentation, because other augmentations such as colour, filter, noise and cutout may affect the desired features of the disease. Diseases, especially those manifesting on the skin or other organs, have specific visual features critical for diagnosis. Color changes, texture, shape, and size of lesions or abnormalities are key identifiers. Augmentations like color, noise, or filter changes might obscure these critical features, leading to a loss of essential diagnostic information. For instance, color augmentations might alter the perceived severity or type of a skin lesion, while excessive noise could mask subtle textural details. Pixel blitting is a technique that involves transferring blocks of pixels from one part of an image to another or modifying them to simulate variations like occlusions or distortions. This process allows for local changes to the image without altering global features, helping preserve the

important visual cues needed for diagnosis while still introducing variability. Geometric augmentations refer to transformations that alter the spatial properties of the entire image or parts of it, such as rotation, scaling, flipping, or cropping. These transformations change the image’s orientation or size but do not directly affect the pixel intensity values like color augmentations would. Geometric transformations help in creating a model that is invariant to the position, orientation, and scale of the disease features, enhancing the model’s ability to recognize these features under various conditions. Together, pixel blitting and geometric augmentations allow for the creation of varied training examples without significantly altering the intrinsic properties of the disease features. They help simulate realistic variations in how diseases might appear in different scenarios without compromising the integrity of the critical diagnostic information. This approach is particularly beneficial in medical/clinical imaging where maintaining accuracy and detail of disease indicators is crucial for effective diagnosis.

Furthermore, Pixel blitting and geometric augmentations offer a more controlled approach to augmenting images, reducing the risk of “leakage” or overfitting during training. Leakage refers to the phenomenon where augmentations become too identifiable, and the model starts recognizing the augmentation itself rather than the underlying patterns it’s supposed to learn. By carefully manipulating the image’s spatial orientation or specific regions via pixel blitting, model architecture maintains a stable and diverse training environment. This stability is crucial for ensuring that the model learns to generalize from the augmented data without becoming overly reliant on augmentation-specific features.

In Transfer Learning setups, where a model is adapted from one task to another, it’s essential to maintain some consistency in how the images are presented. Sudden, drastic changes in image properties can derail the transfer process. Pixel blitting and geometric augmentations provide a gentle yet effective way to augment images. They ensure that the core visual features remain intact and recognizable, aligning with the pre-trained model’s learned features and facilitating a smoother adaptation to the new task.

For instance, in the Transfer Learning set-up, the augmentations were applied too quickly at the early stages of the training. At the very beginning stage of our implementations and setup, a few experiments were carried out with all the given augmentations offered by StyleGAN2-ADA. However, a set of augmentations such as colour, filters, noise and cutout started to leak at the later stages of the training.

One of the augmentations which had shown frequent leaking was the colour augmentation. This problem was also encountered in the work by Karras et al.[194]. Hence those experiments and results are not included in this experimental configurations. The further experiments were set up with a limited set of augmentations and those experiments are listed in Table 3.1.

As in this work a 24 GB GPU was used for the experiments, and several configuration choices required adjustment and recalculation during the experiments. The minibatch size, mini-batch standard deviation, exponential moving average, R_1 regularization γ were altered according to the nature of the input and GPU configuration. The alterations on these hyperparameters are dependent on image resolution and GPU model. The numeric value of these hyperparameters helps in reducing computational space, time, and cost by leading to smoother progress during the training. As the input images resolution 512×512 and the number of GPUs used is 1, the following configurations were used during the training

- the minibatch size = $\max(\min(1 \cdot \min(4096 // 512, 32), 64), 1) = 8$,
- mini-batch standard deviation = $\min(\text{minibatch size} // \text{GPUs}, 4) = 4$,
- Exponential Moving Average = $\text{minibatch size} \cdot 10 / 32 = 2.5$

Among the various implementation choices, R_1 Regularization weight was given utmost importance during the experiments, which will be discussed in further sections.

It is important to measure the quality of image generation for synthetic images. The majority of experiments using StyleGAN2-ADA [194] in the literature have been evaluated using the Frechet Inception Distance (FID). The FID measures the distance between real samples x and generated samples g and is given by:

$$FID(x, g) = \|\mu_x - \mu_g\|^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{0.5}) \quad (3.5)$$

Where:

- μ_x and μ_g are the means of the real and generated samples, respectively.
- Σ_x and Σ_g are the covariances of the real and generated samples, respectively.
- Tr stands for the trace of a matrix.

In this chapter, the experimental results were assessed using Kernel Inception Distance (KID)[265]. The KID is based on the concept of Maximum Mean Discrepancy (MMD) to compute the distance between two distributions. Specifically,

for the context of evaluating GANs, the distance between the distributions of real images P and generated images Q .

Given samples x_1, x_2, \dots, x_m drawn from P and samples y_1, y_2, \dots, y_n drawn from Q , the MMD squared with a certain kernel k is given by:

$$MMD^2(P, Q) = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m k(x_i, x_j) - \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n k(x_i, y_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k(y_i, y_j) \quad (3.6)$$

Where:

- $k(x, y)$ is the kernel function, often chosen as the Radial Basis Function (RBF) or Gaussian kernel:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

with σ as a bandwidth parameter.

The KID is then the empirical estimate of this MMD^2 . A smaller KID value implies that the two distributions (real and generated images) are closer, indicating better performance of the Deep Generative Models.

Lower values of KID indicate better performance. The main reasons to consider KID for the experiments are listed below:

- KID functions outperform FID in case of limited samples i.e., a small number of images.
- KID has a simple, unbiased, and asymptotically normal estimator, in contrast to FID.
- KID compares skewness as well as mean and variance.

As listed in Table 3.1, there are various experimental set-ups explored to obtain high-quality synthetic faces with *Rosacea*. The rationale for chosen parameter values and main findings are outlined below:

- Training from scratch in **Exps 1** and **2** does not provide any advantage with the limited data i.e., 300 input images. However, these experiments show that the γ value has a significant impact in terms of image generation and convergence during the training. As shown in Fig. 3.4, Exp 1 achieved the lowest KID at training step 2640 with $\gamma = 6.5$, whilst Exp 2 achieved the lowest

Table 3.1: List of experiments and results detailing the training setups, augmentation methods, regularization strengths (denoted by γ), and the corresponding performance metrics. The table showcases different training strategies, including training from scratch and Transfer Learning (TL) from the Flickr-Faces-HQ (FFHQ) dataset, with varying degrees of augmentation choices and freeze-D technique. Performance is evaluated using the Kernel Inception Distance (KID) metric \downarrow , with the best KID score achieved noted for each experiment, along with the training step number at which this score was recorded. This comprehensive overview allows for a direct comparison of the impact of training methodologies and augmentation techniques on the quality of synthetic image generation.

Exp no.	Training set-up	Freeze-D	Augmentation Choice	γ	Best KID $\times 10^3$ achieved	At step no.
1	From scratch	NA	blitting, geometry, colour, filter, noise, cutout	6.5	6.8	2640
2	From scratch	NA	blitting, geometry	10	11.8	720
3	TL from FFHQ	NA	blitting, geometry	6.5	3.6	120
4	TL from FFHQ	4	blitting, geometry	6.5	3.5	80
5	TL from FFHQ	13	blitting, geometry	6.5	3.3	680
6	TL from FFHQ	13	blitting, geometry	10	104.6	840
7	TL from FFHQ	13	blitting, geometry	3	3.1	80
8	TL from FFHQ	13	blitting, geometry	2	4.2	360
9	TL from FFHQ	17	blitting, geometry	6.5	3.3	800
10	TL from FFHQ	10	blitting, geometry	6.5	2.5	160

KID at training step 720 with $\gamma = 10$. As shown in the Fig. 3.5(a)(b), The distribution of *Rosacea* artefacts on the generated images from Exp 1 are better compared to the generated images from Exp 2. Hence, it can be concluded that Exp 1 has the best achieved KID and better-quality generated images when training from scratch; conversely, Exp 2 converged faster but generated lower quality images. A lower strength of γ performed better for training from scratch.

- In contrast, Transfer Learning from FFHQ[190] in **Exp 3** performed approximately 33 times better with the improvement in training time/cost and nearly

twice better at the training step 120 with the lowest recorded KID value during the training with the $\gamma=6.5$. As the FFHQ dataset is fundamentally a facial dataset, it was expected to have a wide range of facial features in the resulting generated images. In the Fig. 3.5(c) the generated images have shown a great level of improvement, although image generation quality can be further improved by freezing the top layers of the discriminator to preserve the smaller features of the disease.

- In **Exp 4**, along with Transfer Learning from FFHQ dataset, the *Freeze-Discriminator (Freeze-D)*[266] technique was studied to improve the fine-grained details of *Rosacea* in the synthetic faces. In this experiment, the top 4 layers of the Discriminator were frozen, which improved the result faster, compared to the Transfer Learning without Freeze-D technique. The augmentation choice was kept unchanged to the previous experiment i.e. pixel blitting and geometric transformations. The R_1 regularization weight is set to 6.5. Figure 3.4 represents the obtained KID values during the training process, in which the best value of $KID = 3.5$ is achieved at the step 80. Hence, it is observed that the training process improves relatively faster when the top layers of the discriminator are frozen. As Transfer Learning with Freeze-D presented better results in Fig 3.5(d), that offered motivation to explore various arrangements of Freeze-D.
- Further the Freeze-D technique with Transfer Learning was applied by freezing 13, 10 and 17 layers of Discriminator. In **Exp 5**, the 13 top layers of Discriminator were frozen during the training with the same settings for augmentations i.e., pixel blitting, geometric transformation and $\gamma = 6.5$. The outcome of this experiment is inferior compared to the previous experiment, based on the inconsistency in training and the lowest KID achieved at the later stage of the training i.e. 3.3 is achieved at training step 680. The generated images as shown in Fig. 3.5(e) from this experiment were lower in quality, e.g. most of the facial features are deformed, blurred with leaky background details. To improve this condition, further experiments were carried out with higher and lower strengths of γ while keeping other hyperparameters unchanged.
- Although some higher values of γ were tested while training from scratch in exp 2, they were not used with Transfer Learning, hence $\gamma = 10$ was tested in **Exp 6**. It can be observed from Fig. 3.4 and Table 3.1, that it took

longer to achieve a minimum KID at step 840. The lowest obtained KID in this experiment was the highest KID value recorded among other experiments, proving it the worst KID value recorded. The generated images in Fig. 3.5(f) were highly distorted and unusable in quality. However, it demonstrated the significance of R_1 regularization strength γ . Regardless of training set up, higher values of γ performed worse in terms of convergence and quality of generated images.

- Hence in the next experiments, the lower values of γ were explored. In **Exp 7**, $\gamma = 3$ was examined, while other hyperparameters were kept unchanged as in the previous Exp 6. As was observed in Fig. 3.4, KID drops at the very beginning stage of training i.e. step 80 and then becomes inconsistent. However, this is one of the second lowest KID values achieved among all the experiments resulting in high quality images generated at step 80 with the KID value 3.1. The generated images in Fig. 3.5(g) were with fine-grained details of *Rosacea* and disease patterns and resembled the real-life cases of *Rosacea*.
- To exploit the performance with lower values of γ , **Exp 8** was carried out with $\gamma = 2$. In this experiment, the lowest KID = 4.2 was recorded at training step 360. It was observed from Fig.3.5(h) that the generated samples are deformed at the left bottom portion with the blurred edges. The distribution of the disease feature was inadequate. It was observable that a low value of γ produces such a strong sort of deformity which was not encountered in the previous experiments.
- Furthermore, experiments **Exps 9 and 10** were carried out by freezing 17 and 10 layers respectively with $\gamma = 6.5$ to observe changes due to freezing the layer of Discriminator. Exp 9 shows inconsistency throughout the training process from the beginning. The minimum KID=3.3 is obtained at training step 800. In Fig. 3.5(i), it is observed that the generated images tend to be blurred around the edges and the center. Some samples are negatively affected by the geometric augmentation.
- In **Exp 10**, generated sample images at the best value of KID =2.5 were obtained at the training step no.160. Although Exp 10 has obtained the lowest KID among all the experiments, the generated images are blurred at the edges and center as depicted in Fig.3.5(j). The details of *Rosacea* are absent.

- As the Freeze-D technique with freezing 4, 10, 13, 17 layers of Discriminator were experimented; the results showed that freezing 10 layers helps achieve the lowest value of KID amongst other training setups. However, it is observed that freezing 10 layers leads to too much smoothing which does not help in preserving the details of the disease. Freezing 4,13,17 layers of Discriminator achieved comparatively better results in terms of the value of KID.
- Along with freezing the layers, we have experimented with various strength of R_1 regularization. Adapting various γ values illustrates its significant impact on the training process, the metric (KID) and the generation of synthetic images.
- The impact of γ value can be observed in both settings such as training from scratch and in Transfer learning. Exps 2 and 6 were carried out with higher strength of gamma and they have demonstrated the significance of the value very distinctly. The lower value of γ leads to better results in training, given other implementation choices kept unchanged.
- The choice of R_1 regularization weight/strength γ value depends on the input data. There is a heuristic formula in 3.4 for choosing the numerical value of γ as an initial guess, which calculates the γ value as 6.5. However, tweaking/adjusting this numerical value leads to better results in generating synthetic images with fine-grained details and improved fidelity. It can be acknowledged that the choice of γ value is sensitive when the number of images are short in supply. Lower values of γ perform better compared to the value obtained by applying the heuristic formulae. However, there is a risk in choosing very low values or very high values.

3.4.1 Truncation Trick

The *Truncation trick* was introduced by BigGAN [246]. The truncation trick acts as a boosting strategy for the quality of images. By applying the truncation trick, we can expand the span in the variety of images. The quality of these individual images is comparatively high, and the distribution of disease artefacts is precise. Transforming the images to latent space provides an opportunity to generate 1000 high-quality synthetic images at a time. This is possible with the Truncation trick introduced by BigGAN architecture.

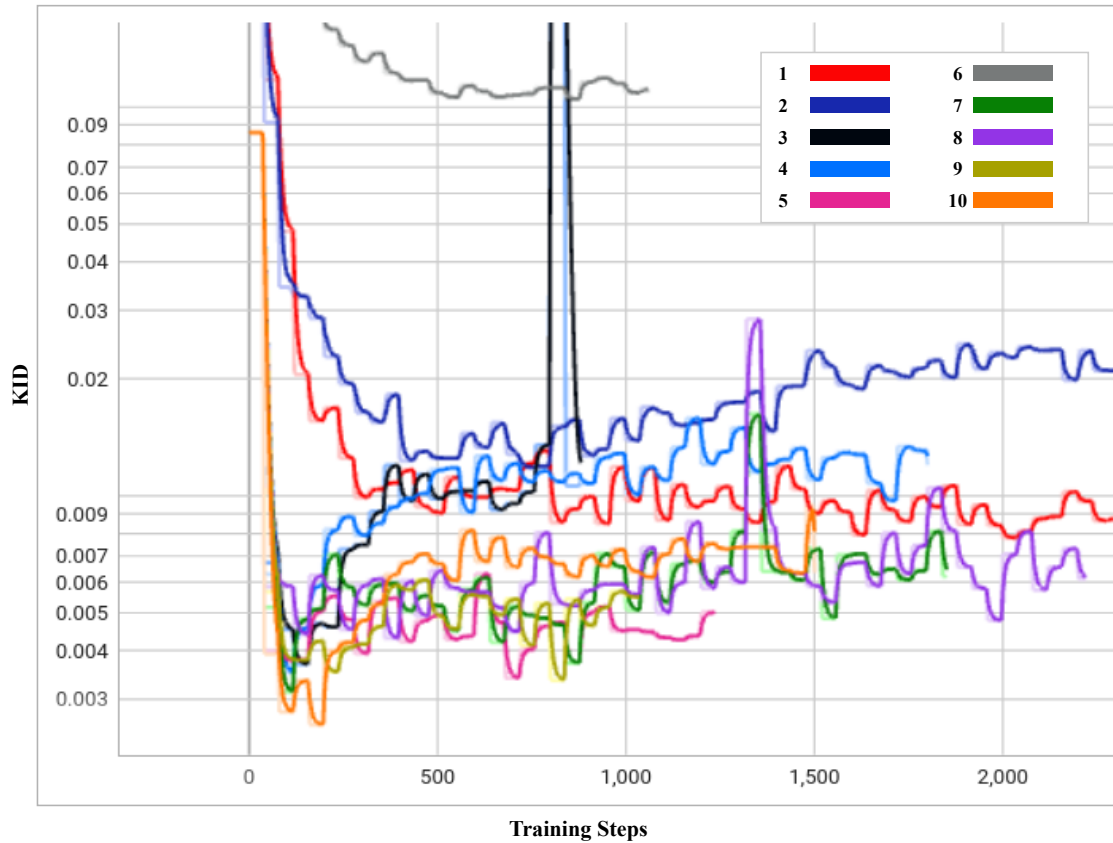


Figure 3.4: Kernel Inception Distance (KID) Progression for 10 Experiments Over Training Period. This graph tracks the KID metric's changes across different training steps for ten distinct experiments. Each coloured line corresponds to an experiment number as indexed in Table 3.1, allowing for a visual comparison of how each experimental setup impacts the KID over time.



Figure 3.5: Generated Faces from Diverse Training Strategies. This figure presents a visual representation of the synthetic faces generated across 10 different experiments as outlined in Table 3.1. Each sub-figure (a-j) corresponds to a unique combination of training setups, augmentation methods, freeze-D and regularization strengths γ detailed in the table. The images showcase the variation in quality and features of the generated faces. This figure, in conjunction with Table 3.1, provides a comprehensive comparison of how each methodology impacts the quality and realism of synthetic image generation.

The truncation trick is a sampling technique that aims at truncating the noise vector z by resampling the values to improve individual sample quality. The truncation trick is regulated by a value called the ‘*truncation threshold*’ ψ . The truncation threshold can lie in the range between 0.5 to 1. As per [43, 246], we have used a truncation threshold ψ of 0.7 to obtain the best observed/most favourable results under the specific conditions and objectives of our study. Choosing the truncation value of 1 indicates that there is no truncation. Different truncation thresholds help in truncating the latent values so that they fall close to the mean. The smaller the truncation threshold, the better the samples will appear in terms of variety.

Although **Exp 10** has achieved the lowest value of KID, the images generated from this experiment are **not useful** due to a few factors made, such as:

- A few images were not properly distributed and they are distorted and blurred with leaked geometric augmentations,
- While exploiting the latent space, most of the samples generated from this experiment lacked variation in regards to common facial features as well as *Rosacea* features,
- As a result, out of 1000 generated images, only 30 high quality images were picked for further analysis.

On the other hand, **Exp 7** achieved the second lowest value of KID, the generated images from this experiment were **useful** due for a few reasons, such as:

- All 1000 sample generated (from training step/epoch 80 with the best KID) were correctly distributed,
- The span of variation was greater than Exp 10, meaning that there was more variety in facial features and *Rosacea* features,
- There were no deformations in the facial and *Rosacea* disease features,
- The samples were not highly smooth in the forehead or cheeks region,
- More distinctive facial and *Rosacea* disease features obtained compared to Exp 10,
- As a result, the best 300 high quality images were picked through visual scrutiny from Exp 7.

Figure 3.6 and Figure 3.7 are the generated images through truncation from Exps 7 and 10 respectively. The images presented in Figure 3.6 and Figure 3.7, the observed blurring of eyes in some images is a direct consequence of the de-identification procedures applied to the underlying real-world dataset of rosacea images. These 300 synthetic images selected from Exp 7 were used for further qualitative analysis discussed in Section 3.5. These 300 images are named as synthetic *rosacea* full faces (**synth-rff-300**) are available on: <https://github.com/thinkercache/synth-rff-300>



Figure 3.6: Generated faces from the best KID value (3.1) of Exp 7 with the truncation $\psi=0.7$



Figure 3.7: Generated faces from the best KID value (2.5) of exp 10 with the truncation $\psi=0.7$

3.5 Qualitative Evaluation of Synthetic Images

Although the best 300 high quality images with resolution 512×512 were selected from the Exp 7, it is important to get them verified by dermatologists to validate the feature and distribution (location/colour/nature) of the *Rosacea*. However, inspecting all the 300 synthetic images is a time-consuming task. Hence out of 300 images, about 50 images were randomly picked for the inspection by the **expert dermatologists**. The images were organised in a Google form. The dermatologists were requested to rate the images from a medical perspective as to how well the artefacts on the generated faces represented *Rosacea* on a linear scale from 1 (not realistic *Rosacea*) to 10 (very realistic *Rosacea*). In total, three dermatologists participated in this research. The scatter plot in Fig. 3.8 illustrates the average rating over the three dermatologists per image. The dots in this 3D plot represent the synthetic images. The darkest colours represent the images with higher ratings followed by the lighter shades for the lower ratings. Fig. 3.9. presents the mean score for each image averaged over the three dermatologists. 73% of the images had a mean score of over 60%.

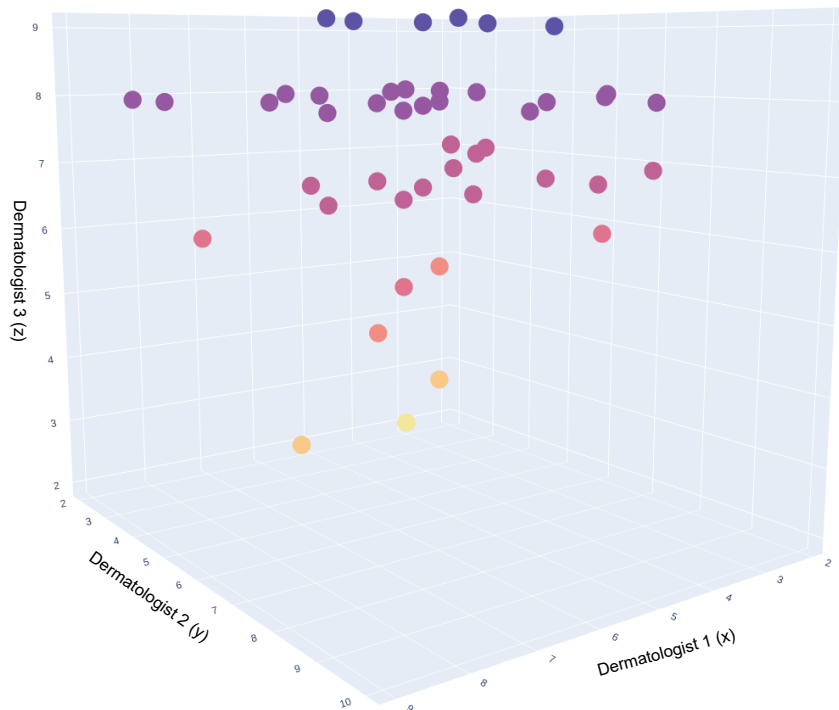


Figure 3.8: A 3D representation of Dermatologists opinion on synthetic images.

Out of 73.07% of images (with more than 60% mean score), 25% of images were rated greater than 80% mean score, 32.7% images were rated greater than 70% to

79% mean score and 15.3% were rated greater than 60% to 69% mean score values as depicted in Fig. 3.10.

To summarize, according to the dermatologists' opinions (in medical perspective), 73% of images present a realistic pattern of *Rosacea* on the generated faces and additional comments provided by the dermatologists are listed on the Table 3.2. Table 3.2 **concludes** that the experts' overall impression of the generated *Rosacea* images is very positive. The feedback from experts reinforces the value of developing synthetic images, highlighting their potential to mitigate the data scarcity issue for *Rosacea* and other facial skin conditions in medical imaging. This positive endorsement from medical professionals underscores the synthetic images' realism and utility, indicating a promising direction for future research and application in the field.

The amalgamation of methodology for synthetic face generation, and from the quantitative and qualitative data, shows an optimistic direction for synthetic data generation for rare skin conditions and other diseases that involves medical imaging. This strategy will help deal with data scarcity problems in many disease domains and facilitate early and faster diagnosis.

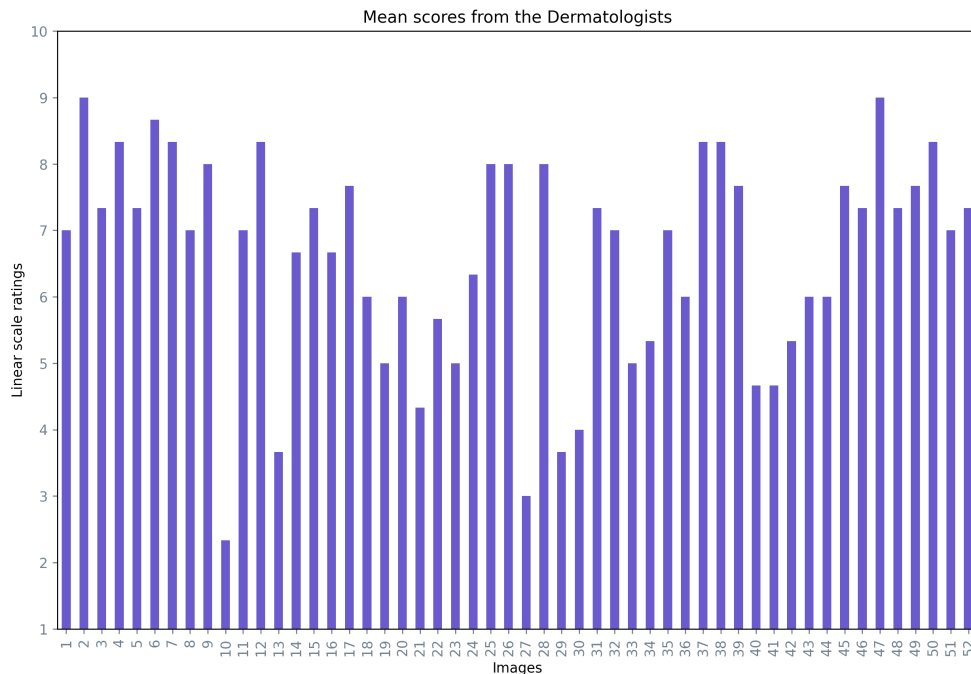


Figure 3.9: Mean scores from the Dermatologists for generated images

The second part of the qualitative evaluation was based on **non-specialist participants'** opinions. In this analysis, a total of 50 images were provided for analysis in which 40 images were generated and 10 images were real. The intention of in-

Table 3.2: Qualitative Evaluation: Dermatologists’ Comments on Generated *Rosacea* Faces. This table presents a collection of feedback and insights from dermatologists regarding the synthetic *Rosacea* faces generated as part of this study. Each entry lists the comments provided by individual dermatologists, reflecting on the realism, potential utility, and overall impression of the generated facial images with *Rosacea* patterns. These comments are instrumental in understanding the clinical relevance and educational value of the synthetic images, as well as guiding future improvements and applications in dermatological training and diagnosis.

Dermatologists	Comments
1	“Diagnosing <i>Rosacea</i> in some patients requires running a lab examination. But, essentially the images in this research created using an artificial intelligence can widely impact the performance of the technologies currently available to dermatologists. I believe these images could also be used for educational purposes if provided with a set of controls to create more variations of the disease. Best of luck.”
2	“I am surprised to see what AI can do. I think this work may help in <i>Rosacea</i> screening later on.” “A few images had a strange form of distortion on the face region but, in general, I am very surprised by the quality of the images and varying intensity of <i>Rosacea</i> in each image.”
3	“Please note, I have only examined the <i>Rosacea</i> and without taking notice of the other characteristics of the faces. I can say ETR is very realistic indeed. Great work, all the best.”

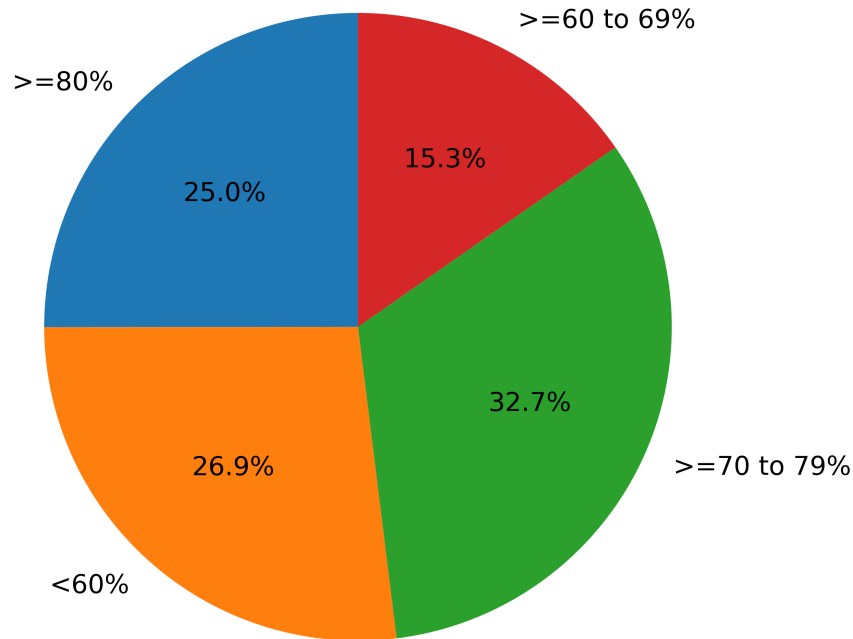


Figure 3.10: Representation of mean scores (in %) for percentage of given images in the study by Dermatologists

cluding 10 real images was to analyse if non-specialist participants could see the difference between the real and fake images. The non-specialist participants were requested to rate the images in the range from 1 (not a realistic face) to 10 (a very realistic face). Fig. 3.11 depicts the mean score range of each image, where generated and real images were labelled in different colours. Out of 50 images, 40 images got the mean score equal to and greater than 60%. Among the top 10 images with highest mean score, 5 images (29, 4, 9, 33, 50) are real and 5 images (6, 3, 5, 1, 23) are generated.

3.6 Limitations and Discussion

Quantitative evaluation of generated images by GAN models particularly in medical imaging is an open-ended problems. Thus, various quantitative and qualitative methods have been adapted and are still in the development stage [267]. The quantitative evaluations are often performed using various metrics such as Inception Score (IS), Fréchet Inception Distance (FID), Kernel Inception Distance (KID), Precision-Recall, and Perceptual Path Length. These metrics are proven to function adequately with certain types of popular datasets which are large in quantity. Although such methods are designed to assess the quality of images or evaluate the

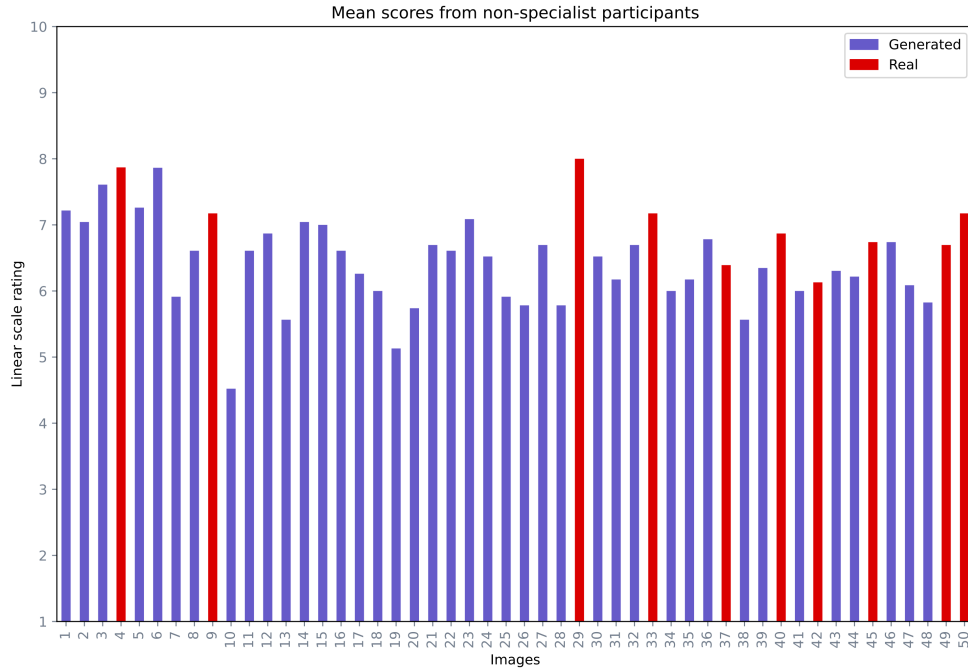


Figure 3.11: Mean scores from the non-specialist participants for miscellaneous images

distribution of the generated images, they may not be a reliable measure for applications in the field of dermatology. These metrics fail to provide any information regarding the “quality” of the generated artifacts on the skin which is vital in diagnosing skin conditions. In the field of dermatology, a minor change on the skin could be meaningful. The existing numerical methods are not capable of measuring the realism of the generated artifacts on the skin and whether they represent a skin condition or not.

As discussed in Section 3.3.3., this research utilises a limited dataset with 300 images to train a generative model. Following the state of the art studies, we deployed a quantitative evaluations pipeline using KID metric to compare the generated images with the real ones. The best value recorded from this evaluation is presented in Table 3.1.

Although Exp 10 achieved the lowest value (the best) in quantitative evaluation with the metric KID, and Exp 7 obtained the second lowest (second best) KID; the images in Exp 7 appear visually more realistic than those in Exp 10. To explore this further, FID metric is calculated, to cross-validate the results by two experiments. The results are reported in Table 3.3. As shown in Table 3.3, the best KID and FID values are obtained from Exp 7 and Exp 10 at different stages of the training process. In Exp 7, the best value obtained by both metrics are at the training step 80; on the

Table 3.3: The FID values are calculated and compared with the top two experiments, selected based on having the lowest KID values.

Exp no.	Top KID value achieved	at step no.	Top FID value achieved	at step no.
7	3.1	80	31.67	80
10	2.5	160	31.40	80

other hand, in Exp 10, the best value obtained by KID is at the training step 160 and the best value obtained by FID is at the step 80. Therefore, it is challenging to measure the realism of *Rosacea* artifact generated on the images based on these quantitative evaluations, specifically for Exp 10. Hence, the images obtained from both experiments needed visual scrutiny to check the fidelity of *Rosacea*.

From visual scrutiny, the generated images from Exp 7 were evaluated of higher fidelity than the images obtained from Exp 10. As discussed in Section 3.4.1, and shown in Figure 3.7, the generated images from Exp 10 are blurred and lack variation in *Rosacea* features. As a result, the images generated from Exp 10 were not included in the further analysis.

As mentioned in Section 3.5, the images obtained from Exp 7 were verified by the experts (dermatologists). Based on to the dermatologists’ opinions, 73% of the images got more than 60% mean score, and the dermatologists remarks are provided in Table 3.2. Based on the non-specialist participants’ opinions, 80% of the images got more than 60% mean score. In a nutshell, the StyleGAN2-ADA with the experimental fine-tuning described earlier in the study produced high-quality realistic results as confirmed by experts and non-specialists participants.

Based on these quantitative and qualitative evaluations, it is conceivable that metrics such as KID and FID are not sufficient by themselves as evaluation criteria when working with a limited dataset of medical images. Both quantitative and qualitative evaluations of the synthetic images demonstrate that, although the evaluation metrics such as FID, IS, and KID are used widely, they have many limitations to be aware of while working with medical images. Along with the quantitative evaluation, the qualitative assessment, such as expert opinion, may well be requisite in the computer-aided medical diagnosis community.

3.7 Future work

Given the importance of the hyperparameter γ as discussed earlier, it would be desirable to design an Adaptive Regularization Technique [268] for the weight matrix to be experimentally tested for StyleGAN2 architectures. Designing an adaptive γ value would not only help in generating high fidelity synthetic images but can help achieve the equilibrium at the early stages during the training with limited samples. Reaching the equilibrium point at the earlier stages may help in reducing training time and cost without compromising the quality in the output.

Adding this adaptive technique for γ may also help in optimizing the model by introducing an automated early stopping point to the training process as it starts to overfit. This may save unnecessary time and cost, while the training is still under progress even after overfitting.

As discussed in section 3.6, popular metrics such as IS, FID, KID, Perpetual Path Length, Precision and Recall should not be considered as the only metrics in the assessment pipeline of synthetic medical images. However, it is necessary to have a quantitative evaluation to navigate the results/outputs by GAN models; hence it is essential to explore and improve the quantitative evaluation methods that may be deemed appropriate for the medical imaging domain. To achieve this, it is crucial to understand the nature of medical imaging with respect to imaging modality, fidelity and how to retain domain-specific information in synthetic images. Addressing the inconsistency between subjective expert evaluations and quantitative metrics like FID and others is a critical challenge, especially prevalent in the medical field. In future, it is necessary to include the development of new algorithms or expert-centric metrics specifically designed to assess the fine-grained details crucial in clinical images. These metrics will aim to capture the subtleties and complexities that are often pivotal in medical diagnosis but may be overlooked by current quantitative methods. By incorporating domain-specific knowledge and diagnostic criteria directly into the evaluation process, these new metrics can provide a more nuanced and clinically relevant assessment of image quality. Furthermore, they could be designed to incorporate direct input from medical experts, potentially using machine learning techniques to learn from expert evaluations and better align with professional standards. The development of such metrics will involve rigorous validation against clinical outcomes and expert opinions to ensure their accuracy and reliability.

The generated images could be used to expand the dataset for the classifica-

tion of *Rosacea* as part of a Deep Learning System (DLS) for skin disease diagnosis [159], which could eventually be utilized in primary care settings [269]. Along with *Rosacea*, additional facial diseases could be included to differentiate the unique features among various facial skin conditions. This method provides an innovative means to address data limitations, thus enhancing the capabilities of these sophisticated systems. Such advanced, deep learning-enabled systems have the potential to augment the decision-making process of physicians by providing corroborative consultations and highlighting areas of concern in clinical/medical images [270]. In future, the application of synthetic images, thereby, might serve as an effective tool in optimizing the diagnostic performance of these systems [242], warranting the need for extensive research in this promising intersection of artificial intelligence and dermatology.

As the quality of the generated synthetic images continues to advance, they could be repurposed beyond their initial intent, proving valuable for increasing *Rosacea* awareness, educational initiatives, and promotional campaigns aimed at disease understanding. Moreover, this methodology could be extended to other facial diseases, thereby broadening its application in dermatological disease awareness and education.

3.8 Conclusion

In this research, we have demonstrated the effectiveness of using StyleGAN2-ADA to generate high-quality synthetic images of *Rosacea* from a small dataset of only 300 real images. By controlling the R_1 regularization weight, we were able to achieve this result, which serves as foundational work for investigating the use of advanced generative models in synthetic data generation for medical imaging with limited data. The conducted experiments also revealed that granular details of the skin disease can be generated by working with hyperparameters such as R_1 regularization, applying a limited set of augmentation techniques such as 'pixel blitting' and 'colour' and the Freeze-D technique with Transfer Learning. A qualitative analysis was conducted, in which expert dermatologists evaluated the generated images of *Rosacea*, and the mean opinion score indicated that 73% of the generated images present a realistic pattern of *Rosacea*. Additionally, this research suggests that metrics such as KID and FID may have limitations in evaluating synthetic images generated from small datasets in the medical and clinical imaging field. The generated images were also

evaluated by non-expert participants, which shows the synthetic *Rosacea* faces look realistic as 80% of the images in the analysis have achieved the mean score 60% and more.

3.9 Data and Code Availability Statement

The input data in these experiments are obtained from 3 sources as following:

1. SD-260 [126]: This dataset has been benchmarked by the study published with the cited reference. The authors Sun et al. [126] have shared the data upon signing the ‘Datasets Request Form’. Hence it is recommended that the interested researchers can access the SD-260 dataset by requesting the first author Xiaoxiao Sun, who kindly shared the dataset with us.
2. Irish Dataset[263, 264]: This dataset, used for our research, has been procured with permission from the Charles Institute of Dermatology, University College Dublin. Researchers interested in accessing this dataset can contact the Charles Institute of Dermatology, University college Dublin <https://www.ucd.ie/charles/>.
3. Images from Google Search results and tele-dermatology websites[112, 113, 114, 75, 117, 119]: The datasets were obtained by performing search queries such as, ‘rosacea subtype 1 ETR rosacea’ and ‘rosacea subtype 2 PPR rosacea’ on Google, as well as looking under the ‘rosacea’ disease section on cited tele-dermatology websites. Only images labelled as ETR and PPR types of rosacea were considered for this research. The data gathering and processing framework was discussed with the Data Protection Unit in Dublin City University and the process is aligned with data protection principles approved by the university.

To support further reproducibility of the work, the code is available: <https://github.com/thinkercache/stylegan2-ada-pytorch>

1. The Exp1-10 experiment configurations (.json) are added to the ‘/Config-Exp1-10’ folder on the <https://github.com/thinkercache/stylegan2-ada-pytorch> repository.
2. The Qualitative Evaluation on Dermatologists and non-specialists participants are shared in the ‘/DermQualitative’ and ‘/NonspecQualitative’ folder

in the <https://github.com/thinkercache/stylegan2-ada-pytorch> repository. These folders contain both qualitative data (.csv) and code (.ipynb).

3. The 300 synthetic *Rosacea* dataset generated in this research/chapter is shared on GitHub repository: <https://github.com/thinkercache/synth-rff-300>.

Chapter 4

3D Modelling of Rhinophyma

A viable method of addressing small classes in the medical imaging domain is to focus on one disease class or subtype at a time. This strategy facilitates the development of specialized models, optimized to make accurate predictions for individual classes of disease or disease subtypes. In addition, it helps balance the dataset by ensuring a sufficient number of samples for each class. By concentrating on specific diseases, we can identify their unique characteristics and tailor our models to better comprehend these features.

This chapter focuses on *Rhinophyma*, a subtype of Rosacea skin condition [14]. *Rhinophyma* is a chronic condition characterized by the gradual enlargement and deformation of the nose. The condition initially appears with mild signs of pore enlargement and acne in the nose region, progressing to moderate symptoms that feature some degree of nose enlargement, and eventually leading to severe cases with significant nose deformities [21]. The details of *Rhinophyma* are discussed in section 4.1. In our research, we have gathered and utilized 268 images of *Rhinophyma*, with further details and information on data acquisition and availability discussed in section 4.2.

Given the limited number of available images for *Rhinophyma*, a challenging but essential skin condition to study, our research focuses on creating sophisticated 3D models of the *Rhinophyma-affected nose*. These models are designed in line with medical literature, reflecting various severity levels of *Rhinophyma* as identified in diagnosis and classification protocols. Our approach involves rendering these 3D models to produce both 2D images and 3D meshes, showcasing a wide range of *Rhinophyma* deformations. The resulting 2D images are intended for two primary purposes. First, they will be used to develop a classification model that aims to

recognize *Rhinophyma* based on the various deformations depicted in the synthetic images. Once the model is developed, its effectiveness will be tested against real-world *Rhinophyma* images to validate its practical utility and accuracy in a clinical setting.

Parametric models are generative in nature, meaning they create or generate new data or models based on the parameters provided. These models are a notably influential practice in 3D modelling. These models leverage a set of parameters and design constraints, thereby facilitating the creation of complex objects that are not only manipulable but also customizable, catering to a diverse array of needs.

Recent advances in synthetic visual data generation, notably 3DGANs [271, 156] and NeRF [272], have highlighted their potency in creating lifelike 3D models. Despite their growing popularity, these methodologies encounter challenges in capturing intricate details with a high degree of control, predictability, and interpretability. Especially for computer vision and graphics applications in the medical field, a well-thought-out design and control of deformations are required, which is not achievable with 3DGAN and NeRF based models. Alternatively, the 3D parametric modelling framework provides significant control over the model by allowing manipulation of individual vertices, edges, and faces to generate specific and detailed features. These models are inherently deterministic and predictable, as the parameters set by the user directly influence model creation. In comparison to generative models, 3D parametric models often offer greater interpretability due to their explicit geometric definitions. Moreover, their ability to render and manipulate in real-time proves essential in areas like surgical planning and tracking. As these models adhere to well-established principles and methodologies for creation and manipulation, they demonstrate significant advantage over advanced generative models such as 3DGAN and NeRF, which require considerable training data to generate high-quality results. This can pose problems, especially when dealing with rare diseases or specific conditions, where sufficient data may be scarce. In such instances, parametric models offer a solution to these limitations, thereby highlighting their indispensable role in visual data generation.

Parametric modelling also offers a significant advantage, in that it facilitates easy alterations to a 3D geometric model's shape by tweaking parameters such as dimensions or curvatures. This approach eliminates the requirement to repeatedly redraw or redesign the model and its elements whenever changes are necessary, leading to considerable time savings. They can be managed through a script that

uses an algorithm to define the dimensions, shape, and various attributes of the 3D object or model [273]. Considering the above advantages, various parametric models are discussed in section 4.3.

Parametric models are incredibly beneficial in medical and clinical imaging fields as they can furnish quantitative information about anatomical structures of interest. These models can assist in the diagnosis and treatment of diseases, planning of surgical procedures, and evaluating the effectiveness of treatments [274]. There are various methods which are used to help with diagnosis and treatment, such as segmentation, registration, and virtual 3D model of the affected part of patients that requires treatment. Medical professionals are increasingly relying on medical imaging such as CT, MRI, ultrasound etc. with additional computer-aided applications, such 3D modelling which has the ability to detect a wide range of anatomical abnormalities. Having access to a 3D model of an anatomical region provides the perspective of realistic height and depth of the deformities. By providing a more personalized and thorough approach to patient care, 3D imaging and modelling have been transforming the medical field [275, 276].

The underrepresentation of certain medical conditions in datasets, a recurring theme in this thesis, results in what is known as a long-tailed distribution. Such distributions are characterized by a few categories containing a large number of instances and many other categories containing very few instances - for example, within our *Rhinophyma* dataset, images belonging to the ‘medium’ severity type are most abundant. Conversely, images representing ‘mild’ severity are less numerous, and those of ‘major’ severity are least frequent, forming the “long tail”. This pattern is a common feature of real-world medical or clinical imaging datasets. To address this imbalance, synthetic data generation can be used to replicate the realistic attributes of diseases that have either manifested or have the potential to, but have yet to be collected or included in existing datasets.

The main purpose of this study is to propose and adopt a 3D modelling pipeline to generate synthetic images of *Rhinophyma* with various possible deformities that are not present in the existing dataset, to capture the 2D images in various angles (perspective) of the face, as well as to generate 3D meshes which can be helpful for further research, development and eventually an aid for healthcare professionals.

In this chapter, we aim to address the data acquisition challenges associated with this disease category, making the following **contributions**:

1. For the first time, synthetic data generation has been demonstrated using only

268 clinical images that cover various stages of *Rhinophyma*.

2. The achievement of synthetic data generation has been facilitated by a parametric 3D face model, designed to encapsulate all possible stages and various patterns of *Rhinophyma*.
3. Through rendering, we generated 2000 possible deformations of a *Rhinophyma* nose. For each of these *Rhinophyma* noses with deformations, images are captured from 10 different perspectives, leading to a total of 20,000 images.
4. To support further research, the synthetic dataset and the related 3D models are made available on an open-source platform.
5. This synthetic *Rhinophyma* dataset has been utilised in the training of a deep learning-based classification model and tested on a real dataset, illustrating the importance of the synthetic dataset through this validation process.

4.1 *Rhinophyma* Diagnosis

Rhinophyma is one of the four subtypes of rosacea [14], a skin condition characterized by various symptoms and signs including transient and non-transient erythema, papules, pustules, and telangiectasias localized on the cheeks, chin, forehead, glabella, around the eyes, and nose. Each subtype of rosacea is associated with distinct skin features, with *Rhinophyma* being linked to the phymatous feature of skin, which presents as benign thickening, surface irregularities, and enlargement, often accompanied by patulous, expressive follicles and telangiectasias; for this reason *Rhinophyma* is also called Phymatous Rosacea. As a late-stage manifestation of rosacea, *Rhinophyma* affects the nasal soft tissues, which results in a gradual breakdown of nasal structure, causing airway obstruction, and deformations of nasal aesthetics [277]. In a nutshell, a *Rhinophyma* affected nose progressively enlarges and becomes distorted [19].

As a common clinical practice, according to el-Azhary et al. [278], the severity of *Rhinophyma* is categorized into three categories such as minor (mild), moderate, major (severe). Later, the *Rhinophyma* Severity Scale (RHISI) was introduced by Wetzig et al. [279]. Currently, the RHISI is a common measure of diagnosis [277]. In this chapter, we consider all the three categories of the el-Azhary [278] scale and scale and all the levels of the RHISI scale[279].

Surgical intervention is necessary for treating advanced *Rhinophyma*, as it involves the removal of excess tissue. The primary objective of surgery is to reduce the overgrown sebaceous glands and reshape the nose, while also promoting the regeneration of the skin on the nose. However, due to the prominent location of *Rhinophyma* on the central part of the face, surgical procedures must be carried out with utmost care [280]. According to the study by Chauhan et al. [277], post-therapy patient satisfaction is frequently reported, regardless of the treatment approach. More than 89% of patients would endorse undergoing *Rhinophyma* treatment, irrespective of the specific method used. Various treatment options exist, with the choice often influenced by both the practitioner’s preference and the patient’s treatment objectives.

3D scanning and modelling is a common approach in diagnosis and treatment for many disease categories [281]. The study by Hollander et al. [282] has discussed the importance of 3D measurements for the surgical treatments in the area around the eye. Similarly, having access to the 3D model of a *Rhinophyma* nose could aid the disease monitoring, and surgical procedures for clinicians.

4.2 *Rhinophyma* Data

Since *Rhinophyma* affects the facial skin, the available dataset for this disease category is limited. Computer-aided diagnosis studies typically rely on tele-dermatology repositories/websites, which also lack *Rhinophyma* datasets. Our previous work[262] lists the most popular teledermatology websites, among which DermnetNZ [75], DermIS [114], dermatoweb.net [283] contain 9, 28, 9 images, respectively. However, SD-260 [126] contains 178 images, which can be obtained upon request from the main authors. In this chapter, 224 images were collected from these sources, but most of the cases are limited to mild and moderate categories of *Rhinophyma*. To supplement the dataset, the additional 44 images were collected from the Google search queries, resulting in a total of 268 images. Nonetheless, the dataset does not encompass every potential deformation caused by *Rhinophyma*. These images served as crucial references and have contributed to the development of the 3D models, discussed in the Section 4.4. Among these, only 220 images are suitable to be used for classification discussed in the Section 4.5. Out of 220 images, there are 77 mild, 118 moderate, and 25 major/severe cases of *Rhinophyma*.

Having access to all possible cases of deformations is crucial for constructing

a statistically accurate computer-aided diagnosis pipeline. This ensures that the dataset has enough diversity and represents the overall population, thus bridging the statistical gap in the data availability.

4.3 Towards 3D Face Modelling Approach

To the best of our knowledge, there is currently no existing work on *Rhinophyma*, specifically in the field of computer-aided analysis of the diseases. However, the extensive review of the literature in Chapter 2 reveals various studies conducted on the broader condition of Rosacea skin. The use of 3D data offers substantial potential in practical applications because it aligns more closely with the physical world, thanks to the additional detailed depiction of object characteristics. This section aims to broadly discuss the implementation and benefits of a parametric approach to 3D modelling and reconstruction, particularly in facial and medical applications.

As discussed in Chapter 2, the parametric modelling frameworks such as Statismo [235, 152] and Scalismo [152] were employed for tasks such as model construction, surface registration, and Active Shape Model fitting. Primarily, Scalismo is devised to generate and analyze statistical models concerning shape and appearance, finding applications in fields like medical imaging and computer vision. However, Scalismo can be demanding in terms of resources, especially during the creation and analysis of large, intricate 3D models. It does not offer extensive options for exporting or converting models into other formats. While Scalismo does provide a handful of pre-existing 3D models that users can alter to generate deformities, it also allows users to export their personalized 3D models. Nonetheless, it does not provide the functionality to construct a 3D face from a single image or a series of 2D images. Furthermore, the basic 3D face model incorporated in the Scalismo software lacks a significant number of vertices for manipulation. During the design of deformations in the nasal region, unintended alterations to other facial features can occur, leading to undesirable stretching of these areas. In situations where extensive vertex manipulation is required, such as in the case of *Rhinophyma* deformities, having increased control over the nose's deformations is crucial. Moreover, Scalismo does not offer direct support for texture mapping or synthesis. However, it can be utilized alongside other software tools for texture mapping and rendering, thus enabling the generation of realistic 3D models with human-like textures. Given these limitations,

Scalismo may not be seen as an ideal solution for our problem.

On the other hand, there are various open-source 3D software options that offer greater flexibility in modelling and can better accommodate deformations with more intricate details. Blender[284] is a popular choice among software options for 3D modelling. It excels in producing high-quality renderings and animations, providing significant flexibility for creative design and experimentation. In addition to its robust graphical capabilities, Blender also supports interactive scripting in Python, which opens avenues for further engineering and customization. Blender is also equipped with a variety of add-on tools or extensions tailored for medical and clinical use, including but not limited to DICOM importer [285], Slicer [286], and Brain3D[287]. For the purpose of our research, we have focused on creating our parametric 3D model from scratch, with the assistance of an artist using Blender. Subsequently, we have modeled the deformities induced by the disease.

4.4 Methodology

4.4.1 3D Models Design Approach

Severity level	Severity score	Description of skin features	Range of severity scores
Mild	0	No evidence of <i>Rhinophyma</i>	0.0 to 0.4
	1	Mild skin thickening	0.5 to 0.9
Moderate	2	Moderate skin thickening	1.0 to 1.4
	3	Strong skin thickening, small lobules	1.5 to 1.9
Major/Severe	4	Lobules with fissures	2.0 to 2.4
	6	Giant <i>Rhinophyma</i>	2.5 to 3.0

Table 4.1: *Rhinophyma* classification by el AZHARY et al. [278], score and range of severity by Wetzig et al. [279] scores (RHISI- Rhinophyma Severity Index) along with skin features as a design reference for 3D modelling.

In the creation of 3D head models, several key elements were taken into account:

1. According to the medical literature, *Rhinophyma* predominantly affects individuals with fair skin of Celtic origin and the Caucasian race, most commonly between the ages between 50 to 70 and *Rhinophyma* skin condition is rare among African American populations and in Asia [277]. Evidence supporting this observation can be found in various datasets, including SD-260 [126],



Figure 4.1: Illustration of 3D head models for both female and male, with severity level 0 (no deformations), representing typical facial features unaffected by *Rhinophyma*.

DermnetNZ [75] from New Zealand, DermIS [114] from Germany, and Dermatoweb.net [283] from Spain. These factors were integral to the decision-making process when determining the 3D design choices. Hence, we have considered these constraints in designing the head models, ensuring that they resemble individuals of Caucasian and Celtic race, specifically within the age group of the 50s. This approach aligns with our objective to create a ‘proof-of-concept’ for the early diagnosis of *Rhinophyma*. The design of the models to mirror individuals in their early 50s is an intentional choice, allowing for the simulation of the particular appearance and texture of the skin commonly observed in this age group.

2. Furthermore, the latest study shows that among 31 people affected by Rhinophyma, there are 30 males for every 1 female [277]. In the existing dataset of 220 images, there are only 5 females compared to 215 males, revealing a pronounced gender disparity in the data acquisition process. Although this imbalance is beyond our control, we can address it by generating synthetic data. To bridge this gender gap, we have taken the initiative to design both male and female head models, a step that aligns with our commitment to equitable representation of gender.

3. The severity of *Rhinophyma*, classified into mild, moderate, and severe (major) categories [278], was an initial consideration. Following this classification, we incorporated all available images and analyzed the deformation patterns associated with *Rhinophyma* to ensure a comprehensive assessment that spans the full spectrum of severity observed in the condition.
4. The data was then organized according to the Rhinophyma Severity Index (RHISI) [279] measure, a diagnostic tool where severity levels are scored from 0 to 6, as outlined in Table 4.1.
5. Images were further arranged based on this severity score, aiding in the precise modelling of nose vertices. To encapsulate all potential real-world deformation features, a numerical range of scores was defined within the design parameters, also detailed in Table 4.1.
6. As shown in Figure 4.1, the deformation is part of a set of variants, labelled as ‘nose variant 1’. In an effort to generate a broad array of deformations, three sets of nose variants were created for both female and male head models. These variants were labelled ‘nose variant 1’, ‘nose variant 2’, and ‘nose variant 3’. Each gender’s nose variants differed from the other, resulting in a total of six unique variants.
7. These nose variants can be adjusted according to the range of severity scores employed in the design (as listed in Table 4.1), with nose variant names selected at random.
8. Given the real-world implications of *Rhinophyma*, it was vital for the 3D models to accurately portray the typical texture of *Rhinophyma*-affected skin. Therefore, textures matching the description of skin features on and around the nasal structure were also incorporated.

4.4.2 *Rhinophyma* 3D Models

Within a 3D space, two entities are defined as follows:

1. Female head model, denoted as X_1
2. Male head model, denoted as X_2

The entities X_1 and X_2 possess a static set of points, termed as C . This immutability is due to the constant overall spatial distribution of each model. Changes to the points in the nose region transpire only when a deformation is applied. The vertices within the nose area are divided into separate sets. These sets, specific to each model, are denoted as a for the female model X_1 and b for the male model X_2 .

For each entity, there exist three nose variant sets. For entity X_1 , these are represented as a_1, a_2, a_3 :

$$a_1, a_2, a_3 \in X_1$$

For entity X_2 , the nose variants are denoted as b_1, b_2, b_3 :

$$b_1, b_2, b_3 \in X_2$$

The quantity of vertex points in each nose variant range $a_1, a_2, a_3, b_1, b_2, b_3$ from 1 to a fixed number of points. The illustration of various 3D head models and various nose variants are illustrated in the Fig 4.2. Both female and male head models embody geometric sophistication with 191,857 vertices, linked together by 383,600 edges. These vertices and edges combine to create 191,744 faces, which are subsequently segmented into 383,488 triangular elements.

4.4.3 Rendering Set-up

The process of rendering involves intricate procedures to create a 2D image or animation from a 3D model or scene. The goal of rendering is to create realistic and visually appealing images that accurately represent the scene or object being rendered. Once the 3D object is created, then the rendering process starts that comprises various series of steps.

The first step of rendering is to position and adjust the 3D model to create the desired composition. Once the model is in place, cameras and lighting are added to the scene to create a realistic appearance. Then rendering settings are applied that include setting the 2D image resolution, choosing the output file format, and camera intrinsics. During or after this process an optional background scene can also be added. Once the rendering starts, the objects/scenes are manipulated according to the prearranged settings which produces animations or 2D images. Similarly, our rendering set-up involves the following steps:

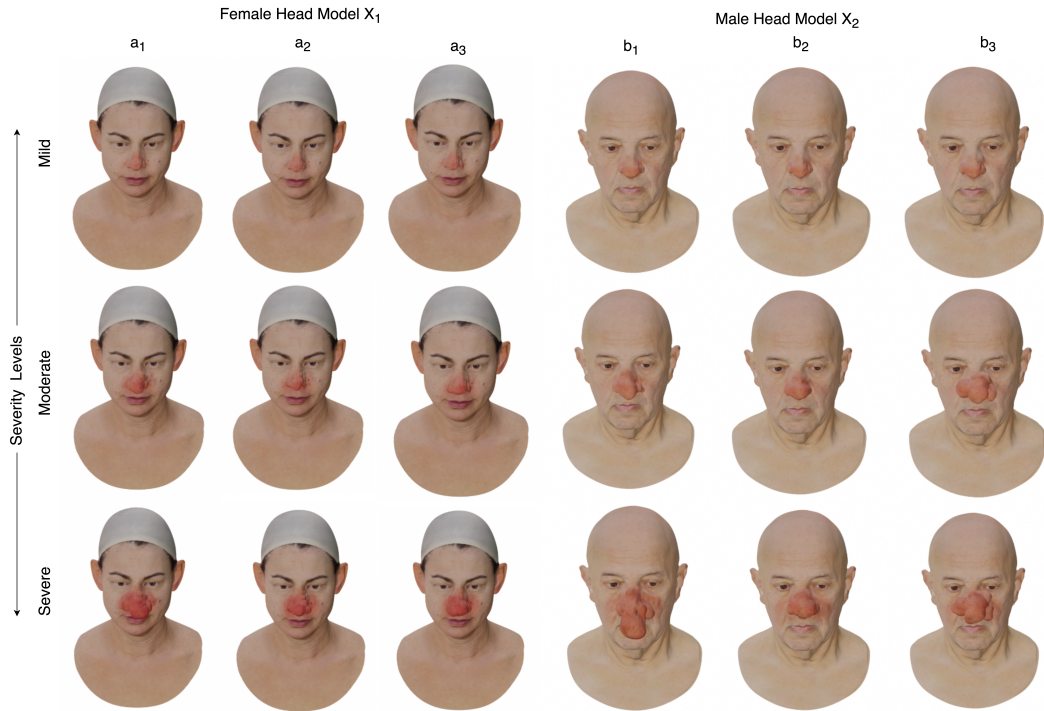


Figure 4.2: An illustration of female and male head models with 6 nose variants at 3 levels of *Rhinophyma* severity.

1. This process commenced with the rendering of two three-dimensional (3D) objects, designated as X_1 and X_2 , utilizing an identical protocol.
2. The rendering process initiates with the placement of object X_1 at the core of the 3D coordinate system, specifically at the coordinates $(0, 0, 0)$. Following its positioning, X_1 undergoes a rotation in Euler mode, precisely 90 degrees around the x-axis, no rotation around the y-axis, and an additional 90 degrees around the z-axis.
3. Each object is constituted by three distinct nose variants (groups of vertices) labeled as a_1 , a_2 , and a_3 . The primary objective was to randomly assign these groups of vertices during each rendering iteration. Moreover, every nose variant had a value range set between 0.5 and 3.0. Values ranging from 0.5 to 1 were classified as mild, values between 1 to 2 as moderate, and those within the range of 2 to 3 as severe.
4. In our experiment, each object underwent the generation of 1,000 random deformations. The selection of the nose variant and the corresponding nose deformation severity level was randomly performed for each deformation. After the generation of *Rhinophyma* nose deformities, a configuration of 10 cameras

was established in a semi-circular arrangement around the object, with a radius of 60 units and a height of 10 units. Each camera was situated at an interval of 18 degrees from the adjacent one. The rationale behind the configuration of ten cameras was to capture the object from varied perspectives, resulting in ten 2D images for each *Rhinophyma* category. This setup is illustrated in Fig. 4.3.

5. Subsequently, the rendering process commenced using Blender’s built-in rendering engine, CYCLES. The output was generated in the form of 2D images of each object, totaling 10,000 images per object and thereby amassing a sum of 20,000 images. The dimension of each output image was set to 960×540 pixels, saved in .png format. Simultaneously, pertinent details related to the rendering process, such as the total number of cameras, active camera name, camera focal length, camera location and rotation details, nose deformation severity, severity label, and active nose variant, were documented in .json format as shown in Fig. ???. The resultant meshes corresponding to each nose deformation severity were archived in .ply format.
6. To optimize the rendering process, each head model underwent rendering in three separate batches to maximize GPU utilization, outputting 350 iterations per model. As the equipped GPU reaches the processing threshold, it exceeds the memory constraints, leading to an automated termination of Blender. Thus, segregating the process into three batches facilitates efficient use of resources while preventing system failures due to memory overflows. The comprehensive rendering process for the male and female head models required approximately 303 hours and 43 minutes, which is equivalent to 12 days, 15 hours, and 43 minutes. The volume of the produced content accumulated to 31.7 GB. The experimental procedures were conducted on a system outfitted with an AMD Ryzen 9 5900X 12 core CPU, an Nvidia Geforce RTX 3060 (12 GB) GPU, and 32 GB RAM, operating on a Linux-based operating system.
7. As a result, we have obtained **395 mild, 809 moderate, and 796 major/severe** cases of *Rhinophyma* in the synthetic dataset.

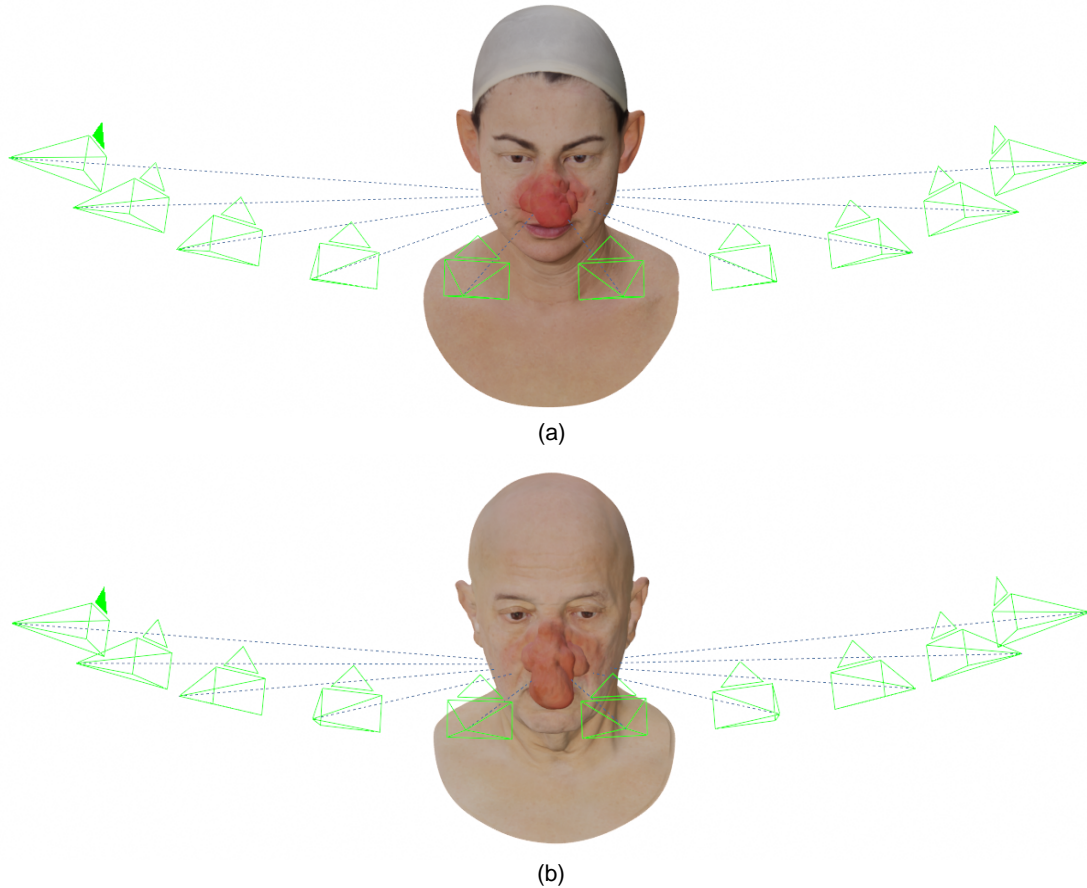


Figure 4.3: An illustration of female and male 3D head models with added deformations that present patterns of *Rhinophyma*. Sub-figure (a) displays the female head model with a deformation applied to nose variant 1 at severity level 2.90. Similarly, sub-figure (b) presents the male head model with a deformation applied to nose variant 1 at the same severity level.

4.5 Synthetic *Rhinophyma* images for Classification

The role of classification is paramount in the realm of computer-aided or deep learning-assisted medical diagnosis. This process enables accurate identification and sorting of diseases or irregularities, thereby guiding the selection of appropriate therapeutic approaches or interventions. Consequently, the verification of synthetically generated images becomes a necessity to confirm their capability to adequately represent features and perform optimally when used with real datasets. The comparative phase of the classification process juxtaposes these synthetically produced *Rhinophyma* images with genuine, normal nose images, which were sourced and readied using the FFHQ dataset [194].

In the context of this research, the binary classification task, distinguishing between *Rhinophyma* and normal nose images, is designed to establish a foundational understanding of the model’s capability to discern distinct features. While binary classification might seem straightforward, its application to medical imaging, particularly with synthetic images, presents unique challenges. These include ensuring the synthetic images are sufficiently detailed and representative of the condition’s variability. Furthermore, in medical diagnosis, even binary classification tasks can be complex due to the subtle and varied manifestations of many conditions. Therefore, the simplicity of the task should not undermine its importance, as it provides an essential step towards more complex, multi-class classification in future studies. Additionally, binary classification serves as a realistic starting point for testing and validating the effectiveness of synthetically generated images before proceeding to more nuanced tasks. It lays the groundwork for more complex analyses and is a common practice in the field to validate new methodologies or technologies.

4.5.1 Synthetic Data Preparation

As discussed in the Section 4.4.1, the head models X_1 and X_2 were rendered through 10 cameras, resulting in 20,000 images. Among these, the frontal view of the faces captured through 3 cameras are considered for the classification, resulting 6000 images (X_1 : 3000 and X_2 : 3000) in total. Nonetheless, as depicted in Fig.4.4, these images incorporate some background elements during the rendering process. In order for a model to learn *Rhinophyma* patterns, the background information is not necessary. Consequently, the face parsing technique, which employs 68 landmark points [288] as illustrated in Fig.4.5, was applied to the synthetically generated *Rhinophyma* images.

The face parsing mechanism follows as below:

- A facial landmark detector was constructed using a pre-trained model from Dlib [289]. This model has the capability to identify the coordinates of 68 critical points on a human face. This method demonstrates its effectiveness particularly when the whole face is visible without any artifacts or obstructions in the facial region.
- Four landmark points were manually identified to form a Region of Interest (ROI) to capture the nose region.



Figure 4.4: Generated images of the female (left) and male(right) images after rendering through the camera no. 5 (on the top), 6 (middle) and 7 (bottom).

- These four points were used to create a mask and extract the aforementioned region surrounding the nose from the rendered images.
- However, this parsing technique could not be applied to some extreme cases of *Rhinophyma* due to the inability to detect the nose and lip regions. Consequently, 37 images remained unparsed, despite attempts with various landmark points.
- Following the application of the optimal parsing technique, the resulting number of synthetic *Rhinophyma* samples is 5963 in total, which are used for classification. The images after applying the parsing are shown in Fig. 4.6.



Figure 4.5: Demonstration of 68-landmark points on the top subfigures illustrated in Fig.4.4

4.5.2 Normal Nose for Classification task

As elucidated in section 4.5.1, the synthetically generated *Rhinophyma* images constitute one class, while the normal nose forms another class, sourced from the FFHQ dataset [194]. The same parsing technique and parameters applied to the synthetic *Rhinophyma* images were also utilized for the FFHQ dataset, resulting in 6000 prepared images for the classification task. The primary objective of this classification process is to differentiate between the normal nose and the *Rhinophyma* nose, necessitating a binary classification. The dataset was partitioned for training, validation, and testing in a ratio of 70:20:10.

4.5.3 Real-world *Rhinophyma* data preparation

As elaborated in Section 4.5.3, a total of 268 authentic/real-life *Rhinophyma* images were amassed from diverse sources, 178 images are from SD-260 dataset [126, 170], and 90 images were gathered from Teledermatology websites [75, 114, 283] and Google Search results.

Of the total 268 images, successful parsing was achieved for 74 images, while parsing was unsuccessful for the remaining 194. The cause of this parsing failure can be traced back to difficulties in facial landmark detection when the full face is not visible or when the eyes appear blurred. Detailed scrutiny of the unparsed 194 images indicated that 48 of them collected from Google search results were unfit for cropping due to the existence of artifacts, such as oversized eyeglasses, reduced resolution, or partial nose visibility. As a result, these 48 images were eliminated, and the remaining 146 images were cropped, focusing on the nasal region.

It is crucial to note that given only partial regions of the faces or noses were visible, the same 48 images were also presented to a 3D artist. The artist was

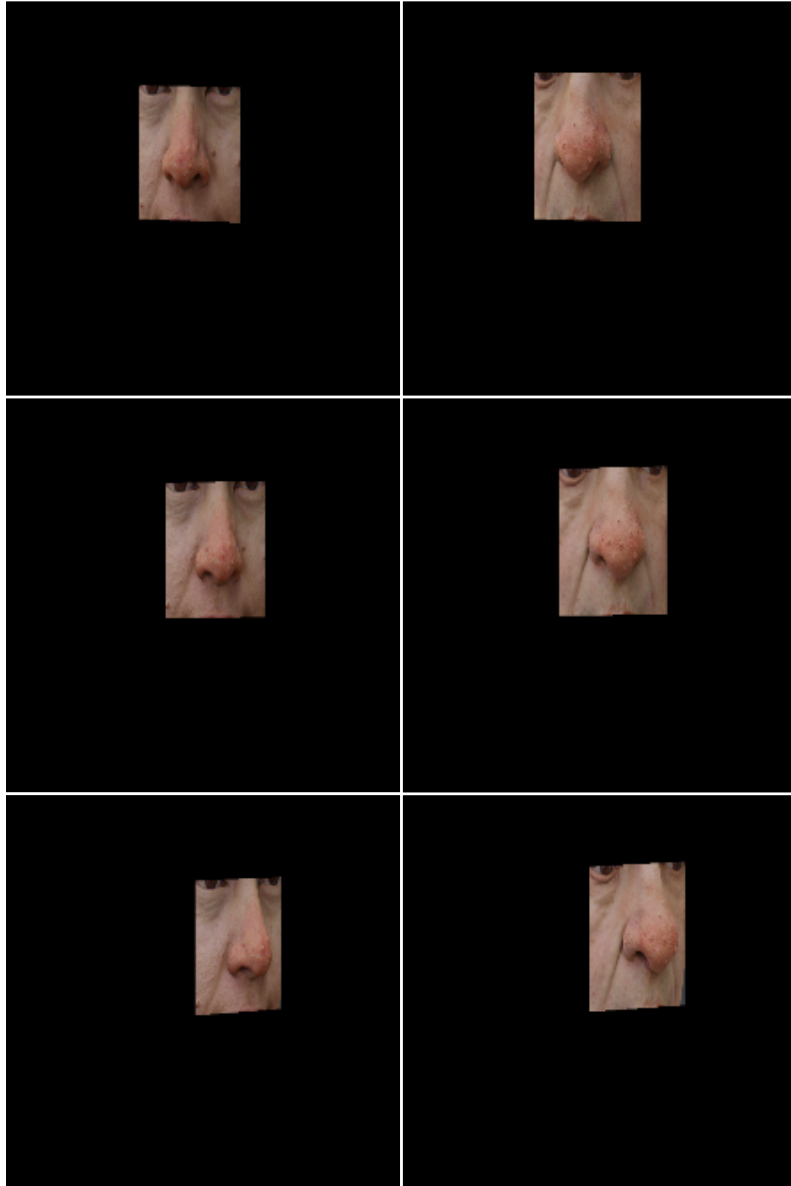


Figure 4.6: Example images of the female (left) and male(right) images through the camera no. 5 (on the top), 6 (middle) and 7 (bottom) after the Parsing technique has been applied.

tasked with understanding and replicating the various stages of *Rhinophyma* (mild, moderate, severe) to accurately create 3D models.

Following this, the 146 cropped images were amalgamated with the 74 images to which parsing was successfully applied, and resized to 224×224 pixels to fit the specifications of the chosen classification models. This yielded a total of 220 processed images, all categorized under the real *Rhinophyma* class. As a result, the most valuable data for the final test set of the real *Rhinophyma* class came from the SD-260 [126, 170] and tele-dermatology websites [75, 114, 283].

To balance this dataset, a supplementary batch of 220 unseen images were prepared from the FFHQ dataset [194], and classified as real normal nose samples. Thus, the final dataset comprises 220 real *Rhinophyma* images labeled ‘rhi’ and 220 real normal nose images labeled ‘norm’. This batch serves as the genuine test set, as it encompasses real-world *Rhinophyma* nose images.

To sum up, the classification task leverages a training set, which includes synthetic *Rhinophyma* images and real nose images, complemented by a validation set and test set constituted of similar categories of images. This is then followed by a test set containing real-world *Rhinophyma* images and real-world normal nose images. Consequently, two test sets are employed: one encompassing synthetic *Rhinophyma* images and the other containing real-world *Rhinophyma* images. The rationale behind the implementation of two separate test sets is to enable a comparative evaluation of the performance of the classification models. This dual test set strategy provides a more robust and comprehensive analysis.

4.5.4 Classification Models

Given the constrained size of our dataset, our exploration is confined to those architectural frameworks that have demonstrated effective performance within resource-limited scenarios such as input data, computational resources, memory, by leading to limited power consumption. The MobileNet family consists of a series of lightweight Convolutional Neural Networks architectures designed and optimized to perform efficiently for limited-resource applications. For this reason, the family of MobileNet is suitable for deployment on mobile and embedded devices. The MobileNet family includes MobileNet-V1 [89], MobileNet-V2 [290], and MobileNet-V3 [291].

MobileNets are efficient neural networks designed for mobile vision applications. The initial version, MobileNet-V1 [89], uses Depthwise Separable Convolutions [292] to create lightweight architectures. It reduces overfitting due to its smaller size and allows fine-grained feature extraction. Knowledge distillation [293] provides facial attribute classification with comparable performance to larger models. MobileNet-V2 [290] surpasses its predecessor in accuracy while maintaining computational efficiency. It introduces inverted residuals and linear bottlenecks that enhance information flow. Innovative features such as linear up sampling and an expansion layer preceding each depthwise convolution enables broader feature extraction and information propagation. MobileNet-V3 [291] employs platform-aware Neural Architecture Search (NAS) for globally optimized network structure. It introduces

Hard-swish, a new activation function, and integrates Squeeze-and-Excitation (SE) [294] blocks into its architecture, enhancing pattern representation. It utilizes efficient layers in its final stage and provides two different variants for different resource constraints and use-cases. In this work, we have experimented on our dataset with MobileNet-V2, and MobileNet-V3-Large.

4.5.5 Results and Discussion

The results using classification models MobileNetV2 [290], MobileNetV3-Large [291] are given in Table 2.

Table 4.2: The classification results. The scores calculated using various metrics on the real test set are highlighted in orange.

Exp.	Model	Test accuracy		Precision		Recall		F1 score	
		Synth	Real	Synth	Real	Synth	Real	Synth	Real
1	MobileNet-V3-Large	1.00	0.89	1.00	1.00	1.00	0.80	1.00	0.89
2	MobileNet-V2	1.00	0.97	1.00	1.00	1.00	0.95	1.00	0.98

Across all three experiments, input images for training, validation, and test sets were subjected to normalization. Augmentation procedures entailing rotation, width and height shifts, zooming, shearing, and both horizontal and vertical flipping were employed.

In our study, we employed transfer learning techniques for both experiments, utilizing pre-trained MobileNet architectures as the foundational base for our models, allowing us to leverage their powerful feature extraction capabilities.

Experiment 1 utilized the pre-trained MobileNet-V3-Large model as a base, enhanced with custom layers: Global AveragePooling2D, two BatchNormalization layers, a Dense layer with 256 neurons (ReLU activation and L2 regularization of 0.01), and a Dropout layer with a rate of 0.5. An output layer was appended, comprising a Dense layer with a single neuron utilizing a sigmoid activation function. Weights of the MobileNet-V3-Large base model were frozen to conserve learned features. An early stopping mechanism, monitoring validation loss with a patience of 3 epochs, was incorporated, alongside a Model Checkpoint to store the best performing model during training. The model underwent training for 20 epochs with a batch size of 32. The model compilation involved binary cross-entropy loss with the Nadam

optimizer, maintaining a learning rate of $1e-5$ and an L2 regularization factor of 0.01.

For Experiment 2, the pre-trained MobileNet-V2 model, supplemented with custom layers: GlobalAveragePooling2D, BatchNormalization, a Dropout layer with a 0.5 rate, a Dense layer with 128 neurons (ReLU activation and L2 regularization of 0.01), and another BatchNormalization layer. The output layer was a Dense layer with a single neuron employing a sigmoid activation function. The MobileNet-V2 base model weights were frozen, allowing only the custom top layers to be updated. Adam optimizer was used with a learning rate of $1e-4$. Early stopping monitored validation loss with a patience of 4 epochs, and Model Checkpoint was used to save the best model. The model underwent training for 50 epochs with a batch size of 32.

Despite the multitude of experiments carried out in this investigation, only the most impactful two results have been presented to maintain conciseness and relevance to the research. The post-training evaluation of our models involves assessments of the evaluation metrics such as test accuracy, precision, recall (sensitivity), F1 score and confusion matrix. Experiment no. 2 performed with the highest test accuracy. Upon thorough assessment, it can be determined that the ‘Recall’ serves as the most effective metric for gauging the model’s performance, as its outcomes align seamlessly with the computations derived from the confusion matrix shown in Table 4.3.

Table 4.3: Confusion matrices for each classification model based on the Second Test Set. The Positive Class is “Normal Nose” and the Negative Class is “*Rhinophyma*”.

Model	Class labels	Confusion Matrix	
		Normal	<i>Rhinophyma</i>
MobileNet-V3-Large	Normal	220 (TP)	0 (FP)
	<i>Rhinophyma</i>	44 (FN)	176 (TN)
MobileNet-V2	Normal	220 (TP)	0 (FP)
	<i>Rhinophyma</i>	10 (FN)	210 (TN)

4.5.6 Class Activation Maps

To delve further into this phenomenon, Gradient-weighted Class Activation Mapping (Grad-CAM) [295] was leveraged as an instrumental tool to reveal the underlying decision-making mechanisms within the classification models. Grad-CAM assists in highlighting the regions within the test image that have the most substantial

influence on the final classification decision. Given that the study was centered around the classification of images conditioned on the presence of *Rhinophyma*, the resulting Grad-CAM heatmaps vividly indicated an emphasis on the nasal regions, as demonstrated in Fig.4.7.

In Fig.4.7, a thorough comparison between the masked and manually cropped images for both models reveals substantial adeptness in picking up the features within the masked category of real-world *Rhinophyma* images, as the classification models were trained on masked images. Intriguingly, the model further displayed its competence in identifying *Rhinophyma* characteristics within manually cropped images, thereby substantiating the efficacy of the image masking approach. This approach encourages the models to concentrate on key features, thus enhancing its capacity to accurately detect real-world instances of *Rhinophyma* in both masked and manually cropped images.

In addition to the quantitative/statistical comparison presented in Table 4.2 and Table4.3, visual comparisons are presented through GradCAM visualizations for both the MobileNet-V3-Large and MobileNet-V2 models, as well as for both masked and manually cropped images. Upon detailed inspection of Fig.4.7, proficiency in identifying features within the nasal anatomy of the face is exhibited by the MobileNet-V3-Large model, independent of whether the images are masked or manually cropped. However, through comparative analysis, it is revealed that the MobileNet-V2 model performs in a superior manner, covering a broader region of the nasal anatomy with a distinctive focus on key features on the masked images. When the performance on manually cropped images is compared, it becomes apparent that MobileNet-V2 excels in localizing its focus on the nasal region, whereas MobileNet-V3-Large demonstrates difficulty in achieving precise localization.

To summarize, while competency in identifying *Rhinophyma* features is displayed by both models, superior performance in feature detection for both masked and unmasked images, particularly localizing *Rhinophyma*, is demonstrated by MobileNet-V2. This observable pattern robustly substantiates the models' ability in accurately identifying salient features of *Rhinophyma*, thereby validating the comprehensiveness of the proposed/implemented approach.

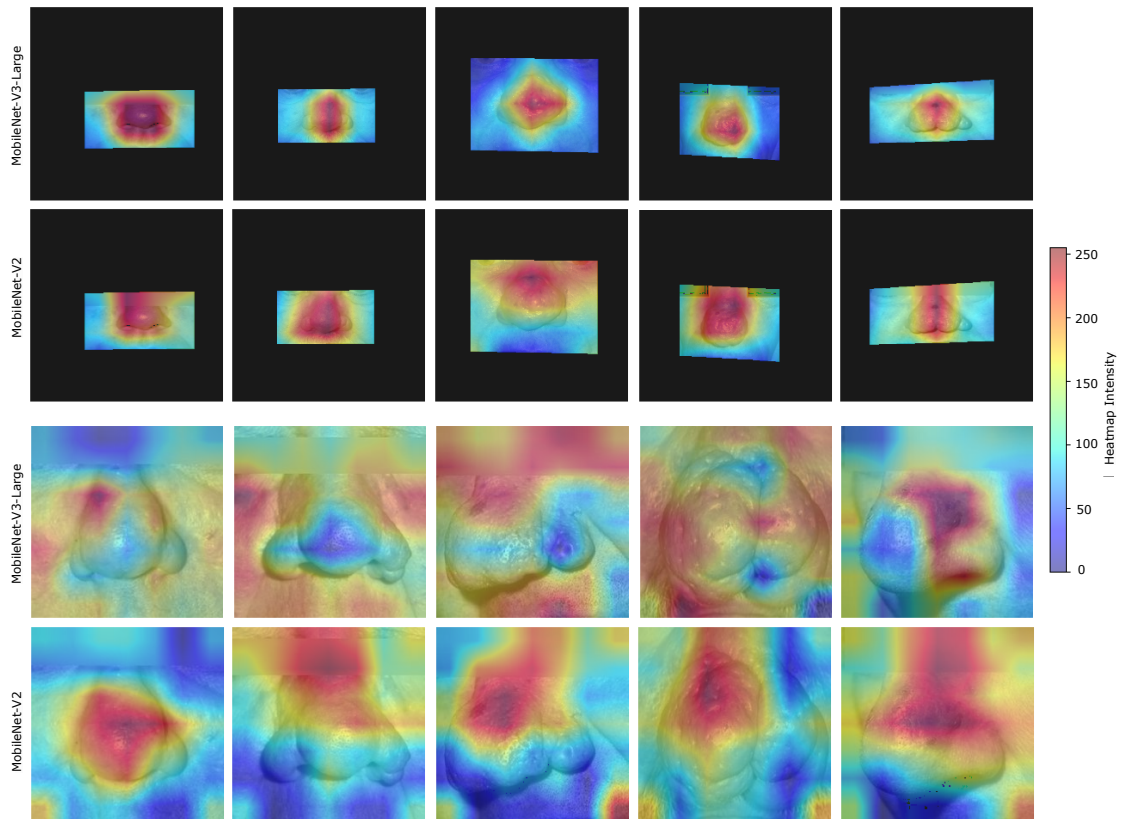


Figure 4.7: Visualisation of feature maps produced by the classification models using GradCAM[295]. Images used in this visualization are taken from SD-260 [126]. The top two rows consist of images that are parsed according to the parsing technique discussed in Section 4.5.1. On the other hand, the bottom two rows feature manually cropped images with an emphasis on the nose region, as discussed in Section 4.5.3.

4.6 Limitations and Discussion

In this chapter, the 3D modelling of Rhinophyma-affected noses and the generation of synthetic data were presented through rendering 3D models for both males and females, resulting in 1000 rhinophyma patterns for each gender. A notable limitation emerged from the scarcity of real-world rhinophyma data for testing the classification models. Within the custom-prepared dataset of 220 real-world Rhinophyma images, only 5 are female, and 215 are male. This gender bias is reflective of what is found in publicly available datasets, such as SD-260 and images from teledermatology websites available for research. According to the latest study on prevalence of Rhinophyma [277], the imbalance may be explained by the observed 1:5 female-to-male ratio affected by Rhinophyma. However, this issue needs careful consideration in future data acquisition for this disease.

The employed masking technique encountered difficulties in identifying synthetic

Rhinophyma instances, especially where notable nasal enlargement toward the lips was present in major or severe cases. These challenges stemmed from the inherent traits of the 68-landmarks algorithm [289], a conventional facial feature recognition method that pinpoints specific features like corners of the eyes, nose tip, and facial contour. For optimal performance, this technique generally necessitates a full or nearly full facial view, and its precision diminishes with obstructed or partial views. In synthetic *Rhinophyma* images with major or severe conditions, the expansion of the nose towards, or overlapping the lips, posed significant detection challenges. Consequently, successful masking of 37 images, including 12 depicting the most severe synthetic *Rhinophyma* cases, could not be achieved. These images, hence, were omitted from the classification task. Alternative algorithms like Haar cascades [296] were also explored, but the standard Haar cascade, typically tailored for full-face contexts, proved unsuitable for nose detection in this instance.

The challenges described above were significantly more pronounced when preparing the real-world *Rhinophyma* dataset for testing the trained classification models. These difficulties were compounded by the inherent photo capturing style, cropping, and anonymization methods used in the SD-260 [126] and teledermatology datasets [75, 114, 283]. These methods concealed other key facial features, focusing on the *Rhinophyma* affected nose. Working with this data proved to be a daunting as well as time-consuming task and the preparation phase necessitated careful manual cropping.

This issue also arose when attempting to mask the background and unnecessary facial anatomy in the FFHQ dataset [194]. The presence of sunglasses and other accessories on the faces in the dataset posed challenges. However, this obstacle actually assisted in filtering the appropriate images for this research, where only the full nose was required to be clearly visible. Since the FFHQ dataset is extensive, accessing and preparing more images with clearly visible noses was not a problem.

Given these challenges, the development of masking algorithms for partially visible patient images is emphasized as crucial. Such algorithms would aid in efficiently masking irrelevant anatomical details and ensure robust anonymization, thereby enhancing the prospects for advanced deep learning-based diagnosis models and expanding the range of potential applications in human face detection and segmentation.

The challenge escalates when images obtained lack adherence to a consistent, systematic protocol. Complications arose during dataset preparation due to the

absence of standardization in clinical digital image capturing; images were often acquired from extreme angles, highlighting the necessity for standardized procedures, attuned to the disease's specific characteristics, to ensure the procurement of qualitative, accurate, and usable data. While this study predominantly aimed at mitigating the medical domain's visual data scarcity through synthetic data, having access to more real-world data could enhance disease understanding. This may enable the generation of higher quality synthetic data in the future, aiding diagnosis and potentially hastening the identification and treatment of Rhinophyma and other facial skin diseases.

4.7 Utility of 3D Modelling and Synthetic Data

Generation of *Rhinophyma*

In this chapter, we have proposed and developed a 'proof of concept' which will benefit the computer vision and medical research communities in the following ways:

- The main goal of this chapter is to generate data for *Rhinophyma* that can help in combatting the data scarcity problem in deep learning models while training for automated classification of *Rhinophyma*. During this study, we have generated 20,000 images of *Rhinophyma* with 2000 possible deformations for 3 levels of severity, and 2000 polygon (.ply) files.
- The generated images and the 3D models can be used for advertising and education purposes while mitigating the hindrance of any privacy and licensing issues while showcasing usage of the images in public.
- The obtained 3D point cloud of the nose can aid in visualising the nasal anatomy in significant details. This visualisation can help in planning the treatment and surgery.
- Followed by the visualisation, the treatment planning can be done with the input of multiple experts for the complex cases, that will help them develop the procedure in a way that will achieve the best possible outcome.
- These models can be manipulated and dissected virtually, providing an effective learning, designing, simulations of surgical procedures, allowing doctors to

practice and refine their planning in a virtual environment before performing surgery on actual patients.

- Further, these can be referred for post-surgery visualisation and record keeping, further consultation and disease condition tracking in the future by enhancing personalized patient care facilities.

4.8 Conclusion and Future Scope

This study has aimed to make contributions towards addressing the data acquisition challenges associated with *Rhinophyma*, a disease category that historically lacked sufficient datasets for robust analysis. This research demonstrates the first successful generation of synthetic data using only 268 clinical images covering various stages of *Rhinophyma*, achieved through specially designed parametric 3D face models. These models facilitated the rendering of 2000 potential deformations of a *Rhinophyma* nose, with each deformation captured from 10 different perspectives, culminating in a dataset of 20,000 images.

The utility and significance of the synthetic *Rhinophyma* dataset was validated through its application in training deep learning-based classification models. Further, testing on a real-world *Rhinophyma* dataset illustrated the value of synthetic *Rhinophyma* data and highlighted the potential of our synthetic data generation method through parametric modelling approach. This method helped in improving the statistical imbalances of the disease, by reducing the long-tailed distribution problem in medical data. This enhances classification accuracy, all of which were possible even with a limited availability of real-world images with long-tailed distribution.

In essence, the innovative approach to synthetic data generation demonstrated in this study offers promising strides towards addressing data scarcity issues. It provides a strong foundation for future research, not only in the study of *Rhinophyma* but potentially for other disease categories as well. It emphasizes the importance of the parametric modelling approach to producing high-quality synthetic data where disease related fine-grained deformations were carefully designed, contributing to more effective disease recognition and treatment strategies.

The developed 3D models can assist in visualizing the nasal anatomy in detail, facilitating treatment planning, surgical simulations, and record keeping for post-surgery consultations and disease tracking. The models offer potential for ma-

nipulation and virtual dissection, providing a safe and effective platform for doctors to refine their treatment strategies.

This study presents a ‘proof-of-concept’ for the computer vision and medical research communities. It also showcases the potential for similar modelling and synthetic data generation approaches for various other diseases that affect anatomical features of the face and other parts of the human body. This can pave the way for advancements in personalized patient care and the development of computer-aided diagnostic models. Furthermore, the results reinforce the efficacy of synthetic data in advancing medical research and the development of more effective deep learning-based diagnostic models, highlighting the potential of such frameworks. Through the combined efforts of 3D artists, computer scientists, and medical experts, methodologies such as 3D parametric models can be significantly valuable in computer-aided medical diagnostic research and application. In an effort to advance further research, these synthetic datasets and corresponding 3D models have been made publicly available.

From a broad perspective, though the creation of 3D models still predominantly requires manual effort, it is important to consider it as a one-time investment. Post this initial investment, it is possible to generate an essentially unlimited volume of impeccably labeled data. This data includes not only RGB images and segmentation maps, but also depth images, stereo pairs from varying viewpoints, point clouds, synthetic video clips, and other modalities.

4.9 Data and Code Availability Statement

The **real-world *Rhinophyma*** data for this study are obtained from the sources as following:

1. SD-260 [126]: This dataset has been benchmarked by the study published with the cited reference. The authors Sun et al. [126] have shared the data upon signing the ‘Datasets Request Form’. Hence it is recommended that the interested researchers can access the SD-260 dataset by requesting the first author Xiaoxiao Sun, who kindly shared the dataset with us.
2. Teledermatology websites: We have obtained the additional data from three teledermatology websites such as DermNet NewZealand (DermnetNZ) [75], Dermatology Information System (DermIS)[114], and Dermatoweb.net [283],

which are available publicly and free to download from respective websites.

The **generated Synthetic Dataset** named “**3D-rhi-synth-2000**” in this study is publicly made available at Zenodo with doi:10.5281/zenodo.8228258 [297].

The **code** and output are made available at: <https://github.com/thinkercache/Rhi-3D-Gen>

Chapter 5

Conclusion

It is important to underscore the significance of research outcomes within both academic and real-world applied research & development contexts. The driving motivation behind this investigation was to explore new research avenues in addressing the data acquisition challenges associated with diseases like *Rosacea and its subtypes*.

The principal goal was to overcome the hurdle of limited data by developing synthetic data generation. To achieve this, we proposed innovative approaches, leveraging Deep Generative Models (specifically, StyleGAN2-ADA) and harnessing parametric modeling concept to custom-designed 3D Face Models.

This chapter delineates the primary methodologies employed, **framing them within the context of the individual Research Questions** posited in this thesis to highlight the contributions made through these outcomes. By revisiting the RQs, we aim to succinctly discuss the rationale behind each methodological choice, the concepts they are anchored in, and the major findings they yielded. Subsequent sections will delve into the identified limitations of this research, paving the way for a discussion on future prospects and a broader outlook.

5.1 Dissertation Overview by Revisiting Research Questions

In Chapter 1, we formulated five research questions, which are addressed in the subsequent chapters (2, 3, and 4) of this thesis. The central motivations behind this research work revolve around exploring strategies to combat **data scarcity** in the medical domain, emphasising **Rosacea**—a prevalent condition notably impacted by

the limited availability of data.

These motivations form the foundation for the principal objective of this study: to confront and navigate the significant challenges posed by data scarcity and the restricted availability of visual data on Rosacea, by providing comprehensive answers to the aforementioned research questions.

5.1.1 Revisiting Research Question- 1

For **RQ1**, which asked, “**What potential approaches exist in current literature to address the issue of limited data for skin disease analysis?**”, we conducted a comprehensive literature review for the first time which is elaborated in Chapter 2 of this thesis. No prior studies deeply explored this theme. Through our research, it was discovered that the fundamental challenge in diagnosing skin diseases lies in the lack of dermatologists in various countries, a fact sourced from numerous departmental survey reports provided by public healthcare entities. The scarcity of dermatologists and the consequent long waiting times underscore the need to broaden the horizon of skin treatments through computer-aided diagnosis. While it is essential to explore approaches for computer-aided diagnoses, another impediment to its widespread adoption is the limited number of images and datasets available for many prevalent skin conditions. Given that machine learning and deep learning models power most modern computer-aided diagnoses, and these models thrive on vast amounts of data, this shortage is concerning. Our examination of 17 publicly available datasets for skin conditions revealed many diseases, some as prevalent as cancer, are underrepresented. Interestingly, while still limited, skin cancer images are more abundant compared to other skin conditions. There are only about 200 images of Rosacea in publicly available datasets. Publicly accessible datasets contain roughly 200 images of Rosacea. Of these, only a few offer a clear full-face view. When contrasted with research based on skin cancer images, the annotated Rosacea images are significantly fewer. This poses a considerable challenge in dividing the dataset for training, validation, and testing of deep learning models.. Additionally, we analyzed several major studies in other medical imaging domains, only to find that datasets are generally scarce for deep learning applications in medical imaging. In this age dominated by ‘big data’, we emphasize the importance of smaller datasets, especially in fields like dermatology that rely on visual inspection for diagnosis.

In addition to these overarching issues, we reviewed numerous studies on Rosacea

and their methodologies. Many Rosacea studies utilized confidential data, making it difficult to obtain and reproduce their results. This limitation further emphasized the importance of leveraging publicly available datasets, regardless of their size. Consequently, we investigated deep learning and computer vision techniques known to yield satisfactory outcomes even with limited data, such as Data Augmentation, Transfer Learning, Generative Adversarial Networks, Meta-Learning, and 3D modeling phenomena.

From our examination of existing studies on data augmentation and transfer learning, several key observations emerge:

- The majority of work using transfer learning for skin disease analysis began around 2016.
- The focus of most studies was on subtypes of skin cancer, such as malignant melanoma and benign nevi.
- While many studies utilized datasets comprising over 1,000 images, only a few worked with datasets smaller than this threshold.
- Although data augmentation and transfer learning techniques have been employed in several studies, they come with certain limitations. While these techniques can transform images by zooming, cropping, flipping, and rotating, they don't always significantly improve results when only a sparse set of data is available for specific skin conditions.
- Implementing these techniques when training a deep learning model requires partitioning the existing dataset into training, validation, and testing subsets. This process can further limit the data's utility, especially in cases where the initial dataset is already small.

To address the challenges of limited data, Generative Adversarial Networks (GANs) have been investigated to produce synthetic data. By generating high-quality, diverse synthetic samples, GANs enable significant augmentation of datasets. This augmentation can alleviate some of the typical constraints faced when dividing limited datasets into separate segments for training, validation, and testing in classification problems. A review of studies centered on GANs reveals:

- The bulk of GAN-related research in the domain of skin disease classification commenced around 2018.

- The diseases predominantly used for creating synthetic image datasets are melanoma or cancerous skin lesions.
- There appears to be no research specifically targeting Rosacea or any other facial skin conditions.
- A minimum of 2,000 real-world images is typically utilized as the starting point for producing synthetic images.
- The ISIC 2017 and 2018 datasets are frequently employed across various studies.
- In terms of architecture, DCGAN, ProGAN/PGAN, and LAPGAN are the most common models applied to generate synthetic images.

For an accurate study of a facial skin condition like Rosacea, there is a pressing need for full-face images that capture detailed anatomy. As a result, the objective for a GAN in this context should be to generate comprehensive facial synthetic images, rather than segments of the face.

Considering contemporary progress in the field, studies by Bissoto et al. [189] and Chai et al. [150] underscore the superior capabilities of noise-based GANs like StyleGAN2 [43] in generating high-quality facial images and distinguishing faces from background noise. These advancements, primarily evaluated using CelebA and FFHQ datasets, suggest promising applications in synthesizing images of facial skin conditions such as Rosacea.

While not a primary focus of this research, we did examine the potential of Meta-Learning and the few-shot classification approach for such applications. However, very limited work has been done in this domain, with most studies beginning around 2020. Insights into the application of this approach for skin diseases are scarce. The motivation to explore Meta-Learning stemmed from its advantages in training and hyperparameter optimization, especially when dealing with limited samples. It can be viewed as an advanced form of classification modeling approach, involving a higher level of optimization. However, given the paucity of studies and insights, we decided not to delve further into this approach within the scope of this thesis. Meta-learning aims to gather knowledge across various tasks to enhance base-level learning and task-specific generalization. While promising, it's still an evolving research area, lacking the maturity of established machine learning methods. Meta-learning models are sensitive to hyperparameters, needing careful tuning, and accurately evaluating

performance in few-shot scenarios is complex, often requiring multiple test runs for reliable results.

We further delved into a variety of 3D face modeling techniques, examining both parametric and statistical models currently in practice. These modeling methods have been a mainstay in medical imaging applications since their inception. Among them, 3D Morphable Face Models and Gaussian Process Morphable Models (GPMMs) stand out as the prevailing approaches for facial modeling. Reasons for choosing 3D modeling for our study include:

- **Extension Beyond Linear Span:** GPMMs can extend beyond the linear span of the training data, making them especially suited for scenarios with limited training data.
- **Generative Nature:** GPMMs are generative, meaning they can capture and reproduce various facial features.
- **Efficient Approximation:** They can approximate shape variations using only a moderate number of leading basis functions or eigenvectors.
- **Precision in Detail:** For many anatomical shapes, detailed deformations often occur only in specific parts of the face. GPMMs excel in capturing these nuanced deformations precisely where they manifest.
- **These modeling methods, while widely used in medical imaging applications, have surprisingly never been employed for modeling facial deformations. This unique gap presented a compelling rationale to the use of 3D modeling for such deformations in our research, especially for a condition like Rhinophyma.**

Intriguingly, this precision holds promise, especially when it comes to integrating expert medical knowledge into the model to accurately represent fine-grained real-world disease deformations. Given that GPMMs combine parametric modeling with a generative approach, we further studied them with the aim of modeling the anatomical changes caused by Rhinophyma. Our goal was to generate synthetic data that encapsulates fine-grained details, closely resembling real-world examples.

In summary, a comprehensive review of existing data on skin diseases, especially Rosacea, coupled with insights from computer vision and deep learning techniques, provides clear guidance for addressing subsequent research questions in the ensuing pages. This in-depth literature exploration offers invaluable insights and charts a

strategic course, enabling the crafting of research methodologies that adeptly confront the challenge of data scarcity through a clearly defined roadmap.

5.1.2 Revisiting Research Question- 2

For **RQ2**, which posed the question “**Can GANs be effectively utilised to generate synthetic images from a limited dataset, enhancing the dataset’s volume and diversity for improved skin disease analysis?**”, we undertook a study to generate high-fidelity synthetic faces for Rosacea subtypes 1 and 2 from limited data, as detailed in Chapter 3 of this thesis. As discussed in our revisit to RQ1, there are approximately 200 images of Rosacea in publicly available datasets/repositories. Of these, only a handful provide a clear full-face view. To address RQ2, we curated a unique dataset from three distinct sources: the Irish Dataset [263], SD-260 [126], and Google search results. This compilation is referred to as “**rff-300 (Rosacea-full-face-300)**”.

In Chapter 3, we delved into prominent works on noise-based GANs, with a focus on synthetic face generation. We scrutinized their strengths and weaknesses as they evolved from the original GANs in 2014 to the state-of-the-art StyleGAN2-ADA in 2021. This examination deepened our comprehension of issues like convergence and mode collapse inherent in these models. Given that StyleGAN2-ADA is designed to address challenges with limited data, it emerged as a suitable candidate for our research question. However, it’s noteworthy that StyleGAN2-ADA was typically trained, developed, and evaluated using a minimum of 1,000 images for experimental validation. In contrast, our study utilized a more constrained dataset of only 300 images, albeit ones with fine-grained and crucial features representing the Rosacea condition. With such a limited dataset, retaining the most salient features during network training and subsequent synthetic image generation becomes challenging. This underscored the need to explore strategies that would complement and enhance the results obtained with the adaptation of StyleGAN2-ADA.

Fundamentally, GANs engage in a Minimax zero-sum game, aiming to learn the data’s distribution—a pattern of pixel values rendering images coherent. The generator’s objective is to craft synthetic images mimicking this distribution so closely that differentiating between real and synthetic visuals becomes a challenge. Given GAN’s foundation on the Zero-sum game principle, the expected outcome is a Nash Equilibrium, where neither component can further optimize without altering the other’s parameters. Central to this is the cost or loss function, influenced by in-

tegrated R_1 Regularization. Achieving minimal divergence between training and model distribution is paramount, targeting the lowest loss at *equilibrium*. It’s recognized that the strength of R1 regularization plays a pivotal role in elevating the quality of generated images. Thus, this study investigates R1 regularization’s impact, aiming to pinpoint the optimal strength of R_1 Regularization γ tailored for the rff-300 dataset.

To synthesize faces representing Rosacea using a dataset of merely 300 images, a series of 10 experiments were executed. Initial experiments incorporated varying experimental settings, encompassing both training from scratch and employing transfer learning from the FFHQ dataset. Notably, the discriminator’s lower layers are oriented towards recognizing generic image features. In contrast, its upper layers are tasked with classifying images based on these features, determining if they are real or generated. The ‘Freezing the Discriminator’ technique, denoted as Freeze-D, offers incremental yet consistent enhancement when combined with ADA. However, on its own, Freeze-D is not potent enough to stave off divergence.

Consequently, the assumption about the parameter gamma became instrumental in the experimental design. Alongside employing the Freeze-D technique, where the discriminator layers were frozen, the impact of different R1 regularization strengths, represented by the parameter γ was investigated. Adjusting the γ values revealed a notable influence on the training dynamics, the KID metric, and the quality of the generated synthetic images. This influence of γ was evident in both experimental conditions: training from scratch and transfer learning. Specifically, experiments utilizing a higher gamma strength unequivocally underscored its importance. Contrary to the value determined by the conventional formula, a smaller γ yielded superior outcomes, particularly when working with limited data sets (in hundreds).

In the course of the experiments, 1000 images were generated from each result, utilizing Truncation Trick[246]. The top-performing experiments, as indicated by the KID metric, were then subjected to qualitative evaluation. From this assessment, Experiment 7 stood out, yielding the highest fidelity images of rosacea. After meticulous visual inspection, 300 high-quality images from this experiment were selected for public release and further validation. These 300 high-fidelity synthetic full-face Rosacea images have been named “**synth-rff-300**” and made available in a public repository for further research and usage by the community.

5.1.3 Revisiting Research Question- 3

In addressing **RQ3**, “**How can synthetic Rosacea images be validated qualitatively by expert dermatologists?**”, approximately 50 images were randomly extracted from the “synth-rff-300” dataset and subjected to qualitative analysis. This was undertaken by three dermatologists and further scrutinized by 23 non-specialist participants.

The rationale behind the dermatologist’s verification was to affirm the synthetic images’ authenticity concerning the features, distribution (location, colour, nature), and overall representation of Rosacea. Yet, while thorough examination is essential, meticulously analyzing all 300 synthetic images was deemed impractical due to the time-intensive nature of such a process. Therefore, a more feasible subset of 50 images was presented to the expert dermatologists for assessment. These professionals were then tasked with rating the synthetic images based on their medical proficiency, gauging the authenticity of Rosacea manifestations depicted therein. They evaluated the images on a linear scale ranging from 1, indicating “not realistic Rosacea”, to 10, symbolizing “very realistic Rosacea”. The cumulative feedback from the dermatologists suggested that 73% of the images showcased a genuine Rosacea pattern. Additionally, specific remarks from the dermatologists further affirmed the positive and authentic representation of the condition in the synthetic images.

The subsequent phase of the qualitative evaluation was steered by the perceptions of non-specialist participants. In this segment, participants were presented with a set of 50 images for assessment. This collection comprised 40 synthetically generated images and 10 real ones. The inclusion of the real images aimed to discern whether non-specialist participants could differentiate between the genuine and the generated photographs. From the evaluation, 40 out of the 50 images received an average score of 60% or higher, signifying that for these 40 images, equivalent to 80%, the representations were perceived as genuine by the participants.

5.1.4 Revisiting Research Question- 4

The fourth research question, **RQ4**, posed the inquiry: “**Is it feasible to generate synthetic data from a limited number of samples using 3D modeling, with precise control over the granular deformations caused by Rhinophyma?**”. To address this, Chapter 4 delves into the use of the parametric modeling approach. This method was employed to model the nasal deformations characteristic of Rhino-

phyma, a subtype-3 manifestation of Rosacea.

The primary objective of this study is to introduce and implement a 3D modeling pipeline. This pipeline aims to generate synthetic images of Rhinophyma that showcase potential deformities absent in existing datasets. Additionally, it captures 2D images from various facial perspectives and creates 3D meshes, benefiting further research, development, and healthcare professionals. Drawing from the literature review in Chapter 2, thorough hands-on experimentation revealed that GPMMs fall short in capturing fine-grained details. Consequently, the decision was made to custom design 3D face models. These models follow the principles of parametric modeling with generative approach and accurately represent Rhinophyma.

To achieve this, Rhinophyma images were first gathered from various sources such as SD-260[126], DermnetNZ [75], DermIS [114], dermatoweb.net [283], and Google search results, totaling 268 images. Out of these, 220 images were used for final testing purposes in the classification models as discussed in 5.1.5, and the rest 48 images served as reference showcasing mild, moderate and severe conditions of Rhinophyma. While these images were not used directly to create 3D models, these images acted as essential references showcasing the real-world appearance of the disease, playing a pivotal role in the development of the 3D models.

In Chapter 4, we delve deeply into the implementation and advantages of the parametric approach to 3D modeling and reconstruction, focusing on its application in facial and medical domains. To the best of our knowledge, no current work exists on Rhinophyma, especially in the domain of computer-aided analysis of the disease. Given that Rhinophyma is a medical condition, it was imperative to embed medical insights derived from well-referenced studies.

Recognizing the demographic prevalence of this condition, our head models were designed to predominantly represent individuals of Caucasian and Celtic descent, specifically in their 50s. Two key aspects were pivotal in designing the 3D models: the Rhinophyma diagnosis classification measures [278] and severity measures [279]. These informed the creation of one female and one male head model, allowing for accurate modeling of nose vertices. To comprehensively capture all conceivable real-world deformation characteristics, a numerical range of scores was integrated into the design parameters. This deformation is associated with a set of vertex group termed ‘nose variant’. To ensure diverse representation of deformations, three distinct sets of nose variants were crafted for both the female and male head models, labeled as ‘nose variant 1’, ‘nose variant 2’, and ‘nose variant 3’. Adjustments to these vertex

groups can be made in alignment with the severity scores employed in the design.

Following the creation of the models, the next step involved rendering them to generate synthetic visual representations of Rhinophyma showcasing a range of deformations. For this purpose, we established 10 camera positions for each of the 3D models, both female and male. Each head model underwent the generation of 1,000 unique deformations. Through the rendering process, we produced a total of 10,000 images for each model. This resulted in an aggregate of 20,000 images, capturing 10 distinct viewpoints for each of the two models across their 2,000 random Rhinophyma deformations. The rendering procedure for both the male and female head models spanned 12 days, 15 hours, and 43 minutes. Within the synthetic dataset, the distribution was such that we obtained **395** mild cases, **809** of moderate severity, and **796** cases classified as major/severe Rhinophyma. The images, the resultant meshes, details related to the rendering process as a dataset have been named “**3D-rhi-synth-2000-Synthetic Rhinophyma Visual Dataset**” and are made available in a public repository on Zenodo.

5.1.5 Revisiting Research Question- 5

To address **RQ5**, which inquires “**Can the classification models trained using synthetic images, derived from a 3D environment perform well when tested on real-world data?**”, we turned to real-world Rhinophyma images to gauge the validity of our synthetic creations. Classification plays an indispensable role in the sphere of computer-aided or deep learning-driven medical diagnosis. Through classification, we achieve precise identification and categorization of diseases or anomalies, subsequently informing the choice of the most fitting treatment methods or interventions. Given this importance, validating the authenticity and reliability of synthetically generated images becomes imperative. It’s essential to ascertain their ability to effectively represent relevant features and deliver satisfactory results when integrated with genuine datasets. For the comparative evaluation in the classification process, our synthetic Rhinophyma images were set against authentic, standard nose images extracted and pre-processed from the FFHQ dataset.

As elaborated in the review of **RQ4**, the 3D head models were rendered, resulting in images captured from 10 different cameras. From these, only the frontal views taken by 3 cameras were deemed suitable for classification, yielding a total of 6,000 images. These images underwent rigorous data processing, which included face parsing utilizing 68-landmark detection[289]. This process emphasized

the Rhinophyma-affected nose while eliminating any extraneous background information inadvertently captured during the rendering phase. For the purpose of classification, the synthetically generated Rhinophyma images were grouped into one class. In contrast, images of normal noses, extracted from the FFHQ dataset[194], formed a separate class. A total of 6,000 images from the FFHQ dataset were subjected to the same data processing methods as the synthetic images. Subsets of the combined datasets of 3D-rhi-synth-2000 and FFHQ were then partitioned into training, validation, and test sets to train the classification models.

Conversely, the 268 images that initially served as reference points in designing the 3D models also underwent several data processing steps to adapt them for classification models. This resulted in a total of **220** processed images, all of which were categorized under the real *Rhinophyma* class. Notably, the most crucial data for the final test set of the real *Rhinophyma* class was sourced from the SD-260 [126, 170] and various tele-dermatology websites [75, 114, 283]. To ensure balance in this dataset, an additional set of 220 unseen images was curated from the FFHQ dataset [194] and designated as real normal nose samples. Therefore, the final dataset consists of 220 real *Rhinophyma* images labeled 'rhi' and 220 real normal nose images labeled 'norm'. This collection is especially significant as the true test set, given its inclusion of authentic *Rhinophyma* nose images.

In summary, the classification task employs a training set comprised of synthetic Rhinophyma images and real nose images. This is complemented by both a validation set and a test set that contain similar categories of images. Subsequently, there's another test set which features real-world Rhinophyma images juxtaposed with real-world normal nose images. This results in the creation of two distinct test sets: one featuring synthetic Rhinophyma images and another with genuine Rhinophyma images. The motivation for employing two discrete test sets is to facilitate a nuanced comparative evaluation of the classification models' performance. This bifurcated test set approach ensures a richer and more comprehensive analysis.

Given the limited size of our dataset, our exploration is restricted to architectural frameworks that have proven to perform well in resource-constrained environments, such as limited input data, computational resources, memory, and energy consumption. The MobileNet family encompasses a range of lightweight Convolutional Neural Networks architectures tailored and optimized for such constrained applications. As a result, we chose MobileNet-V2 [290] and MobileNet-V3 [291] for our experimentation. Through training, MobileNetV2 demonstrated a commend-

able 95% accuracy in detecting Rhinophyma when tested on the dataset comprising real-world Rhinophyma images. To further validate this outcome, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) [295] as a pivotal tool to elucidate the decision-making processes inherent to our classification models.

To summarize the findings from RQ5: although both models show proficiency in identifying Rhinophyma features, MobileNet-V2 demonstrates superior performance in detecting features, especially in localizing Rhinophyma, for both masked and unmasked images. This consistent trend strongly underscores the models' capability to accurately pinpoint the key features of Rhinophyma, thereby affirming the effectiveness and comprehensiveness of the proposed approach.

5.2 Limitations

In this thesis, the primary focus is on overcoming challenges associated with limited data in medical or dermatology cases by implementing various strategies to generate synthetic visual data. This task is particularly challenging since most deep learning models designed for computer vision problems typically require large datasets, whereas our available data comprises only a few hundred samples.

When evaluating the results of the StyleGAN2-ADA experiments quantitatively, we observed that relying solely on quantitative evaluations is not sufficiently reliable. Without the validation and opinions of experts, using synthetic images for medical or clinical purposes becomes questionable. The issue of quantitatively evaluating images generated by GAN models, especially in medical imaging, remains unresolved and is an open-ended challenge. We managed to secure validations from three dermatologists, but obtaining their feedback took approximately two months. This delay is understandable, given that healthcare professionals are often busy. Consequently, securing their time for such evaluations proved to be a time-consuming endeavor. Ideally, the top two experimental results from the GAN would have undergone more thorough scrutiny by dermatologists. However, the extensive time requirements made this impractical. Based on both the quantitative and qualitative evaluations, we concluded that metrics like KID and FID may not be adequate as standalone evaluation criteria, especially when dealing with a limited dataset of medical images.

While developing 3D models for Rhinophyma faces, we encountered several limitations. Within our custom-prepared dataset of 220 real-world Rhinophyma images,

a notable disparity was observed: only 5 images represented females, while a staggering 215 represented males. This gender bias is inherent in the publicly available datasets we utilized for our research. Ideally, a more balanced representation of genders would have potentially offered a more comprehensive design influence. A significant limitation in processing both synthetic and real-world Rhinophyma data was the facial feature detection. Our approach relied on the 68 landmarks algorithm[39], a standard facial feature detection scheme. This algorithm demands a full or near-complete view of the face. Its accuracy wanes in cases of partial or obscured facial views. With synthetic images depicting severe Rhinophyma, the pronounced nasal enlargement, occasionally extending over the lips, impeded accurate detection. Similar detection limitations manifested with real-world Rhinophyma images. Many images lacked a full-view of the patient's face, complicating the processing. Managing this data turned out to be both intricate and time-consuming, requiring painstaking manual cropping for accuracy.

While a considerable amount of historical research has focused on the occurrence of Rosacea in individuals with fair skin, recent studies have begun to explore its manifestation in people with colored skin. These more recent explorations have highlighted distinct variations in clinical presentations, exacerbating factors, potential triggers, and the consequences of Rosacea in individuals with skin of color (SOC) [298, 299, 300].

Even though Rosacea is less frequently reported in SOC, this might be attributed to delayed diagnoses or late presentations. This delay often arises due to the challenge in identifying the classical features of Subtype-1 in darker skin tones [299]. As a result, many individuals with SOC who have Rosacea might experience delayed diagnosis, which can lead to inappropriate or inadequate treatment, increased morbidity, and uncontrolled, progressive disease with disfiguring manifestations, such as Rhinophyma [301].

The majority of images available in the dataset of Rosacea subtypes 1, 2, and 3 predominantly feature individuals with fair skin. Consequently, this predominance influenced the design choices for the 3D models, and colored skin was not incorporated, due to the lack of visual reference materials depicting the manifestation of Rosacea on colored skin.

5.3 Future Work

In this study, we have applied the two-stage validation approach for the generated data. Even though quantitative evaluations are essential to interpret GAN model results, there is an acute need to refine these methods for the medical imaging sector. Achieving this requires a deep understanding of medical imaging's nuances, considering imaging modality, fidelity, and the preservation of domain-specific information in synthetic images.

Any future data collection for Rosacea and its subtypes i.e specially Rhinophyma should consciously strive to address the gender imbalance observed in current datasets. Future studies must consider incorporating diverse skin colors during the data acquisition process, representing a wide range of skin colors and genders. Such inclusivity is crucial as it would not only aid in diagnosing people in specific demographic regions but also render the findings scalable and globally applicable, ensuring a more universally beneficial impact.

The challenges encountered in our study underscore the need for advanced masking algorithms tailored for partially visible patient images. Such solutions can enhance the accuracy of masking unwanted anatomical details, ensuring rigorous anonymization. This progression will not only boost the capabilities of deep learning diagnosis models but will also widen their utility in facial detection and segmentation tasks for partial visible/captured faces, including the ones which were anatomised during the data acquisition.

The dataset preparation highlighted the complications arising from the lack of a standardized approach to clinical image capturing. Many images were sourced from non-standard angles, underscoring the necessity for capturing protocols tailored for diseases like Rhinophyma and other facial skin conditions.

Although our primary focus was addressing data scarcity through synthetic image generation, acquiring more real-world images is indispensable. Having access to the real-world dataset would deepen our understanding of Rhinophyma and other similar conditions. Consequently, this could set the stage for creating higher fidelity synthetic data in the future, elevating the diagnostic procedure and possibly accelerating the recognition and intervention for Rhinophyma and other related facial ailments.

The efficacy of state-of-the-art deep learning models is inherently dependent on the volume and quality of data available for training. Yet, there remains ambiguity around the ideal volume of data required for these models to achieve optimal per-

formance. This ambiguity is further complicated by factors such as the diversity of image features and the specificities of the dataset in question. As such, emphasizing the collection of a diverse range of disease features becomes imperative, particularly when considering various disease categories and their corresponding subtypes.

Through the course of this research, a clear distinction emerged within the sphere of dermatology. While skin cancer diagnoses are anchored in the medical imaging domain, primarily because related conditions are captured using dermatoscopes, other skin ailments do not necessarily adhere to this convention. Often, these conditions are documented using conventional digital photography in clinical settings, rather than specialized medical imaging equipment. This distinction underscores the need to standardize, or perhaps introduce, a domain specifically focused on these clinical images. By doing so, the quality and specificity of data in this segment could be elevated, leading to more robust and accurate diagnostic models in the future.

As we look toward expanding the capabilities of computer-aided diagnosis in dermatology, a particularly promising avenue involves the adaptation of healthy human face images to include features of rosacea. Neural style transfer [302] could be adapted to apply the ‘style’ or characteristic appearance of Rosacea onto images of faces without the condition, effectively blending the texture and color patterns associated with the disease onto otherwise healthy-looking faces. Concurrently, attribute manipulation techniques (image-to-image-based GAN models) such as facial attribute editing GAN (AttGAN) [303] or Selective Transfer Network GAN (STGAN) [304] provide a focused approach, allowing specific features or regions of the face to be altered to simulate Rosacea, ensuring that changes are both localized and realistic. Semantic image synthesis could further refine this process by using semantic segmentation maps to guide the model on where rosacea features should appear, ensuring an accurate and contextually appropriate application of the condition’s visual markers. Lastly, Deep Feature Interpolation (DFI) [305, 306] offers a subtler method, enabling a gradual transition between healthy and affected skin by interpolating between the deep features of images with and without Rosacea. By integrating these methods, a robust and nuanced approach can be developed, offering realistic and varied representations of Rosacea on human faces, enhancing both the understanding of the disease’s visual impact and the development of more effective diagnostic tools.

Looking ahead, the utilization of foundation models, especially large-scale Transformer models [307], presents an exciting frontier for computer vision in medical

diagnosis. These models have already demonstrated remarkable success in various domains of artificial intelligence due to their ability to learn rich, transferrable features from large and diverse datasets. In the medical/clinical imaging sector, such models could revolutionize the way we understand and analyze complex patterns in clinical data, offering enhanced accuracy, adaptability, and efficiency. Future work will explore the integration of these powerful models into Rosacea and skin disease diagnostic frameworks, focusing on customizing and fine-tuning them to recognize and classify a wide range of skin conditions with high precision. This will involve not only technical adaptations of the models to suit specific medical imaging tasks but also rigorous validation against clinical outcomes to ensure their reliability and effectiveness in real-world settings. Moreover, as these models require substantial computational resources and data, we will investigate strategies for efficient training and deployment, ensuring they are accessible and practical for medical professionals.

The rise of vision-based transformer [308] models in computer vision holds promising implications for medical diagnosis, particularly in enhancing generalization capabilities across traditional computer vision tasks, it is acknowledged that their application is still at an early stage. These models treat image patches as sequential data, similar to how original transformers process words, allowing them to capture intricate patterns and relationships within medical images. Their ability to understand both the local and global context of an image enables a more comprehensive and nuanced feature representation, crucial for identifying subtle anomalies in medical diagnostics. Furthermore, vision transformers are scalable and exhibit improved performance with larger datasets, a significant advantage given the growing size of medical image repositories. The inherent flexibility of these models in handling various input sizes and their reduced reliance on extensive data augmentation can streamline the diagnostic process, making it more efficient and less prone to errors, limited features and lack of standardization as we observe in medical images. Notably, the attention mechanisms integral to transformers offer a level of interpretability that is vital in clinical settings, providing insights into the model's decision-making process and highlighting critical areas in the images.

In addition to enhancing traditional computer vision tasks, the adaptation of vision-based foundation models also opens avenues for multimodal integration, combining semantic and geometric understanding with other modalities such as natural language. This integration is particularly evident in applications like visual question answering, image captioning, and instruction following, where the model leverages

both visual and textual data to perform complex tasks. In the realm of medical imaging, integrating clinical NLP-related information can provide a more comprehensive understanding of patient data. For instance, combining diagnostic images with patient histories or clinical notes could enable models to learn better patterns and provide more accurate diagnoses. Luo et al. [309] discusses unimodal models typically struggle to achieve high accuracy in dermatology due to the vast variability in skin appearances across different patients and are further hampered by privacy concerns and the shortage of adequately labeled data. However, recent developments in Transformer architectures and large-scale pre-training models based on unlabeled data offer new opportunities for improving AI diagnostic models. Multimodal neural networks, such as CLIP [308], that leverage both text and image information can potentially increase diagnostic accuracy significantly with the integration of Federated Learning. As these models continue to evolve, their integration into medical imaging workflows promises to enhance diagnostic accuracy, aid in treatment planning, and ultimately improve patient outcomes, marking a significant leap forward in the application of AI in healthcare.

Bibliography

- [1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [2] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.
- [3] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [4] Andre Esteva et al. “Deep learning-enabled medical computer vision”. In: *NPJ digital medicine* 4.1 (2021), p. 5.
- [5] Gaël Varoquaux and Veronika Cheplygina. “Machine learning for medical imaging: methodological failures and recommendations for the future”. In: *NPJ digital medicine* 5.1 (2022), p. 48.
- [6] Mingyu Kim et al. “Deep learning in medical imaging”. In: *Neurospine* 16.4 (2019), p. 657.
- [7] Junghwan Cho et al. “How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?” In: *arXiv preprint arXiv:1511.06348* (2015).
- [8] Riccardo Miotto et al. “Deep learning for healthcare: review, opportunities and challenges”. In: *Briefings in bioinformatics* 19.6 (2018), pp. 1236–1246.
- [9] Justin M Johnson and Taghi M Khoshgoftaar. “Survey on deep learning with class imbalance”. In: *Journal of Big Data* 6.1 (2019), pp. 1–54.

- [10] Zeju Li, Konstantinos Kamnitsas, and Ben Glocker. “Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*. Springer. 2019, pp. 402–410.
- [11] L. Checklist. *AI Data Engineering*. Cognilytica. 2020.
- [12] Sergey I Nikolenko. *Synthetic data for deep learning*. Vol. 174. Springer, 2021.
- [13] M Augustin et al. *European Dermatology Health Care Survey 2013 Short Report*. 2013.
- [14] Frank C Powell. “Rosacea”. In: *New England Journal of Medicine* 352.8 (2005), pp. 793–803.
- [15] James Q Del Rosso et al. “Why is rosacea considered to be an inflammatory disorder? The primary role, clinical relevance, and therapeutic correlations of abnormal innate immune response in rosacea-prone skin.” In: *Journal of drugs in dermatology: JDD* 11.6 (2012), pp. 694–700.
- [16] Martin Steinhoff, Jürgen Schaubert, and James J Leyden. “New insights into rosacea pathophysiology: a review of recent findings”. In: *Journal of the American Academy of Dermatology* 69.6 (2013), S15–S26.
- [17] L Gether et al. “Incidence and prevalence of rosacea: a systematic review and meta-analysis”. In: *British Journal of Dermatology* 179.2 (2018), pp. 282–289.
- [18] SA Johnston et al. “Experiences of rosacea and its treatment: an interpretative phenomenological analysis”. In: *British Journal of Dermatology* 178.1 (2018), pp. 154–160.
- [19] Frank Powell. *Rosacea: diagnosis and management*. CRC Press, 2008.
- [20] PJ Hampton et al. “British Association of Dermatologists guidelines for the management of people with rosacea 2021”. In: *British Journal of Dermatology* 185.4 (2021), pp. 725–735.

- [21] Yalçın Tüzün et al. “Rosacea and rhinophyma”. In: *Clinics in dermatology* 32.1 (2014), pp. 35–46.
- [22] National Rosacea Society. *Red Skin & Rashes Are Not Always The Result Of Rosacea*. URL: <https://www.rosacea.org/blog/2016/june/red-skin-rashes-are-not-always-the-result-of-rosacea>.
- [23] Dermatologytimes. *2021 Rosacea Report*. 2021. URL: <https://www.dermatologytimes.com/view/top-20-trending-stories-of-2020>.
- [24] USnewshealth. *Treatments for Rosacea*. 2021. URL: <https://health.usnews.com/health-care/patient-advice/articles/treatments-for-rosacea>.
- [25] David A Huffman. “Impossible objects as nonsense sentences”. In: *Machine intelligence* 6 (1971), pp. 295–323.
- [26] Maxwell B Clowes. “On seeing things”. In: *Artificial intelligence* 2.1 (1971), pp. 79–116.
- [27] Hans P Moravec. “The Stanford cart and the CMU rover”. In: *Proceedings of the IEEE* 71.7 (1983), pp. 872–884.
- [28] Timothy F Cootes et al. “Active shape models-their training and application”. In: *Computer vision and image understanding* 61.1 (1995), pp. 38–59.
- [29] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. “Active appearance models”. In: *Computer Vision—ECCV’98: 5th European Conference on Computer Vision Freiburg, Germany, June 2–6, 1998 Proceedings, Volume II* 5. Springer. 1998, pp. 484–498.
- [30] Timothy F. Cootes, Gareth J. Edwards, and Christopher J Taylor. “Active appearance models”. In: *IEEE Transactions on pattern analysis and machine intelligence* 23.6 (2001), pp. 681–685.
- [31] Volker Blanz and Thomas Vetter. “A morphable model for the synthesis of 3D faces”. In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 1999, pp. 187–194.

- [32] Bernhard Egger et al. “3d morphable face models—past, present, and future”. In: *ACM Transactions on Graphics (TOG)* 39.5 (2020), pp. 1–38.
- [33] Yann LeCun et al. “Handwritten digit recognition with a back-propagation network”. In: *Advances in neural information processing systems* 2 (1989).
- [34] Lorien Pratt. “Reuse of neural networks through transfer”. In: *Connection Science (Print)* 8.2 (1996).
- [35] Jeremy West, Dan Ventura, and Sean Warnick. “Spring research presentation: A theoretical foundation for inductive transfer”. In: *Brigham Young University, College of Physical and Mathematical Sciences* 1.08 (2007).
- [36] Andrew Ng. “Nuts and bolts of building AI applications using Deep Learning”. In: *NIPS Keynote Talk* (2016).
- [37] Yann LeCun et al. “Learning algorithms for classification: A comparison on handwritten digit recognition”. In: *Neural networks: the statistical mechanics perspective* 261.276 (1995), p. 2.
- [38] Geoffrey E Hinton et al. “Improving neural networks by preventing co-adaptation of feature detectors”. In: *arXiv preprint arXiv:1207.0580* (2012).
- [39] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).
- [40] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in neural information processing systems* 27 (2014).
- [41] Anton Osokin et al. “GANs for biological image synthesis”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2233–2242.
- [42] Ashish Shrivastava et al. “Learning from simulated and unsupervised images through adversarial training”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2107–2116.

- [43] Tero Karras et al. “Analyzing and improving the image quality of stylegan”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 8110–8119.
- [44] Amirata Ghorbani et al. “Dermgan: Synthetic generation of clinical skin images with pathology”. In: *Machine learning for health workshop*. PMLR. 2020, pp. 155–170.
- [45] Jingkuan Song et al. “AgeGAN++: Face aging and rejuvenation with dual conditional GANs”. In: *IEEE Transactions on Multimedia* 24 (2021), pp. 791–804.
- [46] Guillaume Le Moing et al. “Semantic palette: Guiding scene generation with class proportions”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 9342–9350.
- [47] Danfeng Hong et al. “Multimodal GANs: Toward crossmodal hyperspectral–multispectral image segmentation”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.6 (2020), pp. 5103–5113.
- [48] Tomáš Hodaň et al. “Photorealistic image synthesis for object instance detection”. In: *2019 IEEE international conference on image processing (ICIP)*. IEEE. 2019, pp. 66–70.
- [49] Weichao Qiu et al. “Unrealcv: Virtual worlds for computer vision”. In: *Proceedings of the 25th ACM international conference on Multimedia*. 2017, pp. 1221–1224.
- [50] Yue Yao et al. “Simulating content consistent vehicle datasets with attribute descent”. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer. 2020, pp. 775–791.
- [51] David Acuna, Jonah Philion, and Sanja Fidler. “Towards optimal strategies for training self-driving perception models in simulation”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 1686–1699.

- [52] Adrien Gaidon et al. “Virtual worlds as proxy for multi-object tracking analysis”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4340–4349.
- [53] Amlan Kar et al. “Meta-sim: Learning to generate synthetic datasets”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 4551–4560.
- [54] Stephan R Richter et al. “Playing for data: Ground truth from computer games”. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*. Springer. 2016, pp. 102–118.
- [55] German Ros et al. “The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 3234–3243.
- [56] Xavier Puig et al. “Virtualhome: Simulating household activities via programs”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8494–8502.
- [57] Samin Khan et al. “ProcSy: Procedural Synthetic Dataset Generation Towards Influence Factor Studies Of Semantic Segmentation Networks.” In: *CVPR workshops*. Vol. 3. 2019, p. 4.
- [58] Xingchao Peng et al. “Visda: The visual domain adaptation challenge”. In: *arXiv preprint arXiv:1710.06924* (2017).
- [59] Matthew Johnson-Roberson et al. “Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?” In: *arXiv preprint arXiv:1610.01983* (2016).
- [60] Fisher Yu et al. “Bdd100k: A diverse driving dataset for heterogeneous multitask learning”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 2636–2645.

- [61] Erroll Wood et al. “Fake it till you make it: face analysis in the wild using synthetic data alone”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 3681–3691.
- [62] Shubhajit Basak et al. “C3I-SynFace: A synthetic head pose and facial depth dataset using seed virtual human models.” In: *Data in Brief* 48 (2023), p. 109087.
- [63] Roderick J Hay et al. “The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions”. In: *Journal of Investigative Dermatology* 134.6 (2014), pp. 1527–1534.
- [64] Divya Seth et al. “Global burden of skin disease: inequities and innovations”. In: *Current dermatology reports* 6.3 (2017), pp. 204–210.
- [65] C Flohr and R Hay. *Putting the burden of skin diseases on the global map*. 2021.
- [66] Matthias Augustin et al. *European Dermatology Health Care Survey 2013 Short Report*. 2013. URL: <https://www.dermasurvey.eu/wp-content/uploads/eu-derma-health-care-survey-2013-short.pdf>.
- [67] British Association of Dermatologists and The King’s Fund. *How can dermatology services meet current and future patient needs while ensuring that quality of care is not compromised and that access is equitable across the UK?* 2014. URL: <https://www.bad.org.uk/shared/get-file.ashx?id=2347%5C&itemtype=document>.
- [68] HSE Dermatologists Ireland. *Dermatology Report*. 2014. URL: <https://www.hse.ie/eng/staff/leadership-education-development/met/plan/specialty-specific-reviews/dermatology-2014.pdf>.
- [69] P Régine Mydlarski et al. “Dermatologic Training and Practice in Canada: An In-Depth Review”. In: *Journal of Cutaneous Medicine and Surgery* 24.3 (2020), pp. 297–303.

- [70] Alex M Glazer et al. “Analysis of trends in geographic distribution and density of US dermatologists”. In: *JAMA dermatology* 153.4 (2017), pp. 322–325.
- [71] Department of Health Australian Government. *Dermatology 2016 Facesheet*. 2016. URL: <https://hwd.health.gov.au/webapi/customer/documents/factsheets/2016/%20Dermatology.pdf>.
- [72] Chang-Bing Shen et al. “Assessment of imaging diagnosis ability of skin tumors in Chinese dermatologists”. In: *Chinese medical journal* 132.17 (2019), p. 2119.
- [73] Paulina Pala, Beata S Bergler-Czop, and Jakub M Gwiżdż. “Teledermatology: idea, benefits and risks of modern age—a systematic review based on melanoma”. In: *Advances in Dermatology and Allergology/Postpy Dermatologii i Alergologii* 37.2 (2020), p. 159.
- [74] Francesc X Marin-Gomez et al. “Diagnosis of skin lesions using photographs taken with a mobile phone: an online survey of primary care physicians”. In: *Journal of Primary Care & Community Health* 11 (2020), p. 2150132720937831.
- [75] New Zealand Dermatological Society. *DermNetNZ*. URL: <https://dermnetnz.org/>.
- [76] Acne and Rosacea Society of Canada. *The National Rosacea Society has conducted surveys of rosacea sufferers to offer insights into the effect on social life*. URL: <https://www.rosaceahelp.ca/impact/social-life/>.
- [77] J Spoenclin et al. “A study on the epidemiology of rosacea in the UK”. In: *British journal of dermatology* 167.3 (2012), pp. 598–605.
- [78] Acne and Rosacea Society of Canada. *Rosacea affects more than 3 million Canadians*. URL: <https://www.rosaceahelp.ca/>.
- [79] M Berg and S Lidén. “Postmenopausal female rosacea patients are more disposed to react with migraine”. In: *Dermatology* 193.1 (1996), pp. 73–74.

- [80] Grand View Research. *Rosacea Treatment Market Size, Share & Trends Analysis Report By Drug Class (Alpha Agonists, Antibiotics, Retinoids, Corticosteroids), By Mode of Administration (Topical, Oral), By Region, And Segment Forecasts, 2019 - 2025, Report ID:GVR-2-68038-739-1*. 2019. URL: <https://www.grandviewresearch.com/industry-analysis/rosacea-treatment-market>.
- [81] TheIrishTimes. *Rosacea: What is it and how can you manage it?* 2021. URL: <https://www.irishtimes.com/life-and-style/fashion/beauty/rosacea-what-is-it-and-how-can-you-manage-it-1.4575944>.
- [82] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *nature* 542.7639 (2017), pp. 115–118.
- [83] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.
- [84] Kenneth Thomsen et al. “Deep learning for diagnostic binary classification of multiple-lesion skin diseases”. In: *Frontiers in medicine* 7 (2020), p. 604.
- [85] Zhixiang Zhao et al. “A Novel Convolutional Neural Network for the Diagnosis and Classification of Rosacea: Usability Study”. In: *JMIR medical informatics* 9.3 (2021), e23415.
- [86] Haijing Wu et al. “A deep learning, image based approach for automated diagnosis for inflammatory skin diseases”. In: *Annals of translational medicine* 8.9 (2020).
- [87] Chen-Yu Zhu et al. “A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment”. In: *Frontiers in medicine* 8 (2021).
- [88] Evgin Goceri. “Diagnosis of skin diseases in the era of deep learning and mobile technology”. In: *Computers in Biology and Medicine* 134 (2021), p. 104458.

- [89] Andrew G Howard et al. “Mobilenets: Efficient convolutional neural networks for mobile vision applications”. In: *arXiv preprint arXiv:1704.04861* (2017).
- [90] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
- [91] Evgin Goceri. “Deep learning based classification of facial dermatological disorders”. In: *Computers in Biology and Medicine* 128 (2021), p. 104118.
- [92] Gao Huang et al. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.
- [93] Christian Szegedy et al. “Inception-v4, inception-resnet and the impact of residual connections on learning”. In: *Thirty-first AAAI conference on artificial intelligence*. 2017.
- [94] Mingxing Tan and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6105–6114.
- [95] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.
- [96] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [97] Mary L McHugh. “Interrater reliability: the kappa statistic”. In: *Biochemia medica* 22.3 (2012), pp. 276–282.
- [98] 1st Lt Pushkar Aggarwal. “Data augmentation in dermatology image recognition using machine learning”. In: *Skin Research and Technology* 25.6 (2019), pp. 815–820.

- [99] Hamidullah Binol et al. “Ros-NET: A deep convolutional neural network for automatic identification of rosacea lesions”. In: *Skin Research and Technology* 26.3 (2020), pp. 413–421.
- [100] Bin Xie et al. “XiangyaDerm: a clinical image dataset of asian race for skin disease aided diagnosis”. In: *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*. Springer, 2019, pp. 22–31.
- [101] François Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.
- [102] Yann LeCun. “The MNIST database of handwritten digits”. In: <http://yann.lecun.com/exdb/mnist/> (1998).
- [103] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- [104] Daniel Shu Wei Ting et al. “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes”. In: *Jama* 318.22 (2017), pp. 2211–2223.
- [105] Robert Lindsey et al. “Deep neural network improves fracture detection by clinicians”. In: *Proceedings of the National Academy of Sciences* 115.45 (2018), pp. 11591–11596.
- [106] Yifan Peng et al. “DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs”. In: *Ophthalmology* 126.4 (2019), pp. 565–575.
- [107] Olivier Bernard et al. “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?” In: *IEEE transactions on medical imaging* 37.11 (2018), pp. 2514–2525.

- [108] Abdur R Feyjie et al. “Semi-supervised few-shot learning for medical image segmentation”. In: *arXiv preprint arXiv:2003.08462* (2020).
- [109] Zhen Zhao et al. “Ensemble Model with Batch Spectral Regularization and Data Blending for Cross-Domain Few-Shot Learning with Unlabeled Data”. In: *arXiv preprint arXiv:2006.04323* (2020).
- [110] Jeremy Kawahara et al. “Seven-point checklist and skin lesion classification using multitask multimodal neural nets”. In: *IEEE journal of biomedical and health informatics* 23.2 (2018), pp. 538–546.
- [111] Seung Seog Han et al. “Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm”. In: *Journal of Investigative Dermatology* 138.7 (2018), pp. 1529–1538.
- [112] Dermatology Atlas Brazil. *Dermatology Atlas Brazil*. URL: <http://www.atlasdermatologico.com.br/index.jsf>.
- [113] An Atlas of Clinical Dermatology Denmark. *An Atlas of Clinical Dermatology*. URL: <https://danderm.dk/atlas/index.html>.
- [114] DermIS. *DermIS*. URL: <https://www.dermis.net/dermisroot/en/home/index.htm>.
- [115] DermNet Skin Disease Atlas. *DermNet Skin Disease Atlas*. URL: <http://www.dermnet.com>.
- [116] University of Edinburgh. *Dermofit Image Library*. URL: <https://licensing.edinburgh-innovations.ed.ac.uk/i/software/dermofit-image-library.html>.
- [117] Dermatoweb.net. *Dermato web spain*. URL: <http://dermatoweb.net>.
- [118] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”. In: *Scientific data* 5.1 (2018), pp. 1–9.

- [119] Hellenic Dermatological Atlas. *Hellenic Dermatological Atlas*. URL: <http://www.hellenicdermatlas.com/en/>.
- [120] Veronica Rotemberg et al. “A patient-centric dataset of images and meta-data for identifying melanomas using clinical context”. In: *Scientific data* 8.1 (2021), pp. 1–8.
- [121] ISIC Archive. *ISIC Archive*. URL: <https://www.isic-archive.com>.
- [122] Ioannis Giotis et al. “MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images”. In: *Expert systems with applications* 42.19 (2015), pp. 6578–6585.
- [123] Molemap NewZealand. *Molemap NewZealand*. URL: <https://www.molemap.co.nz>.
- [124] Zongyuan Ge et al. “Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging”. In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE. 2017, pp. 986–990.
- [125] Teresa Mendonça et al. “PH 2-A dermoscopic image database for research and benchmarking”. In: *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2013, pp. 5437–5440.
- [126] Xiaoxiao Sun et al. “A benchmark for automatic visual classification of clinical skin disease images”. In: *European Conference on Computer Vision*. Springer. 2016, pp. 206–222.
- [127] SD198. *SD198*. URL: <http://xiaopingwu.cn/assets/projects/sd-198/>.
- [128] Ammara Masood and Adel Ali Al-Jumaily. “Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms”. In: *International journal of biomedical imaging* 2013 (2013).

- [129] Barbara Rosado et al. “Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis”. In: *Archives of Dermatology* 139.3 (2003), pp. 361–367.
- [130] Marco Burroni et al. “Melanoma computer-aided diagnosis: reliability and feasibility study”. In: *Clinical cancer research* 10.6 (2004), pp. 1881–1886.
- [131] Harold Kittler et al. “Diagnostic accuracy of dermoscopy”. In: *The lancet oncology* 3.3 (2002), pp. 159–165.
- [132] David Gutman et al. “Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)”. In: *arXiv preprint arXiv:1605.01397* (2016).
- [133] M Binder et al. “Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network.” In: *Melanoma research* 8.3 (1998), pp. 261–266.
- [134] Shunichi Amari et al. *The handbook of brain theory and neural networks*. MIT press, 2003.
- [135] Bradley J Erickson et al. “Machine learning for medical imaging”. In: *Radiographics* 37.2 (2017), pp. 505–515.
- [136] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique”. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1153–1159.
- [137] Maryellen L Giger. “Machine learning in medical imaging”. In: *Journal of the American College of Radiology* 15.3 (2018), pp. 512–520.
- [138] June-Goo Lee et al. “Deep learning in medical imaging: general overview”. In: *Korean journal of radiology* 18.4 (2017), pp. 570–584.
- [139] Kenji Suzuki. “Overview of deep learning in medical imaging”. In: *Radiological physics and technology* 10.3 (2017), pp. 257–273.

- [140] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.
- [141] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115.3 (2015), pp. 211–252.
- [142] Nicolas Papernot et al. “The limitations of deep learning in adversarial settings”. In: *2016 IEEE European symposium on security and privacy (EuroS&P)*. IEEE. 2016, pp. 372–387.
- [143] Ian Goodfellow. “Nips 2016 tutorial: Generative adversarial networks”. In: *arXiv preprint arXiv:1701.00160* (2016).
- [144] Lan Lan et al. “Generative adversarial networks and its applications in biomedical informatics”. In: *Frontiers in Public Health* 8 (2020), p. 164.
- [145] Salome Kazeminia et al. “GANs for medical image analysis”. In: *Artificial Intelligence in Medicine* (2020), p. 101938.
- [146] Lilian Weng. “Meta-learning: Learning to learn fast”. In: *Lil’Log <https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html>* (2018).
- [147] Chelsea B Finn. *Learning to learn with gradients*. University of California, Berkeley, 2018.
- [148] Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks”. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.
- [149] Tim Cootes, ER Baldock, and J Graham. “An introduction to active shape models”. In: *Image processing and analysis* 328 (2000), pp. 223–248.
- [150] Tim F Cootes and Christopher J Taylor. “Statistical models of appearance for medical image analysis and computer vision”. In: *Medical Imaging 2001: Image Processing*. Vol. 4322. International Society for Optics and Photonics. 2001, pp. 236–248.

- [151] David Cristinacce and Timothy F Cootes. “Boosted regression active shape models.” In: *BMVC*. Vol. 2. Citeseer. 2007, pp. 880–889.
- [152] Marcel Lüthi et al. “Gaussian process morphable models”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.8 (2017), pp. 1860–1873.
- [153] Aaron S Jackson et al. “Large pose 3D face reconstruction from a single image via direct volumetric CNN regression”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 1031–1039.
- [154] Anh Tuan Tran et al. “Regressing robust and discriminative 3D morphable models with a very deep neural network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5163–5172.
- [155] Xian-Feng Han, Hamid Laga, and Mohammed Bennamoun. “Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.5 (2019), pp. 1578–1604.
- [156] Baris Gecer et al. “Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 1155–1164.
- [157] Connor Shorten and Taghi M Khoshgoftaar. “A survey on image data augmentation for deep learning”. In: *Journal of Big Data* 6.1 (2019), pp. 1–48.
- [158] Sinno Jialin Pan and Qiang Yang. “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.
- [159] Yuan Liu et al. “A deep learning system for differential diagnosis of skin diseases”. In: *Nature medicine* 26.6 (2020), pp. 900–908.
- [160] Evgin Goceri. “Skin disease diagnosis from photographs using deep learning”. In: *ECCOMAS thematic conference on computational vision and medical image processing*. Springer. 2019, pp. 239–246.

- [161] Md Ashraful Alam Milton. “Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018: Skin lesion analysis towards melanoma detection challenge”. In: *arXiv preprint arXiv:1901.10802* (2019).
- [162] Xiaoyu Cui et al. “Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review”. In: *Journal of the American Academy of Dermatology* 81.5 (2019), pp. 1176–1180.
- [163] Chanki Yu et al. “Acral melanoma detection using a convolutional neural network for dermoscopy images”. In: *PloS one* 13.3 (2018), e0193321.
- [164] Arkadiusz Kwasigroch, Agnieszka Mikołajczyk, and Michał Grochowski. “Deep neural networks approach to skin lesions classification—A comparative analysis”. In: *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE. 2017, pp. 1069–1074.
- [165] Adria Romero Lopez et al. “Skin lesion classification from dermoscopic images using deep learning techniques”. In: *2017 13th IASTED international conference on biomedical engineering (BioMed)*. IEEE. 2017, pp. 49–54.
- [166] Sara Hosseinzadeh Kassani and Peyman Hosseinzadeh Kassani. “A comparative study of deep learning architectures on melanoma detection”. In: *Tissue and Cell* 58 (2019), pp. 76–83.
- [167] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol. “A scoping review of transfer learning research on medical image analysis using ImageNet”. In: *Computers in biology and medicine* 128 (2021), p. 104115.
- [168] Zirui Wang et al. “Characterizing and avoiding negative transfer”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 11293–11302.
- [169] Sourav Mishra, Toshihiko Yamasaki, and Hideaki Imaizumi. “Supervised classification of Dermatological diseases by Deep learning”. In: *arXiv preprint arXiv:1802.03752* (2018).

- [170] Jufeng Yang et al. “Clinical skin lesion diagnosis using representations inspired by dermatologist criteria”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 1258–1266.
- [171] Wannipa Sae-Lim, Wiphada Wettayaprasit, and Pattara Aiyarak. “Convolutional neural networks using mobilenet for skin lesion classification”. In: *2019 16th international joint conference on computer science and software engineering (JCSSE)*. IEEE. 2019, pp. 242–247.
- [172] Kemal Polat and Kaan Onur Koc. “Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all”. In: *Journal of Artificial Intelligence and Systems* 2.1 (2020), pp. 80–97.
- [173] Khalid M Hosny, Mohamed A Kassem, and Mohamed M Foad. “Classification of skin lesions using transfer learning and augmentation with Alex-net”. In: *PloS one* 14.5 (2019), e0217293.
- [174] Amirreza Mahbod et al. “Skin lesion classification using hybrid deep neural networks”. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1229–1233.
- [175] Danilo Barros Mendes and Nilton Correia da Silva. “Skin lesions classification using convolutional neural networks in clinical images”. In: *arXiv preprint arXiv:1812.02316* (2018).
- [176] Zhiwei Qin et al. “A GAN-based image synthesis method for skin lesion classification”. In: *Computer Methods and Programs in Biomedicine* 195 (2020), p. 105568.
- [177] Haroon Rashid, M Asjid Tanveer, and Hassan Aqeel Khan. “Skin lesion classification using GAN based data augmentation”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 916–919.

- [178] Alec Radford, Luke Metz, and Soumith Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks”. In: *arXiv preprint arXiv:1511.06434* (2015).
- [179] Emily Denton et al. “Deep generative image models using a laplacian pyramid of adversarial networks”. In: *arXiv preprint arXiv:1506.05751* (2015).
- [180] Christoph Baur, Shadi Albarqouni, and Nassir Navab. “MelanoGANs: high resolution skin lesion synthesis with GANs”. In: *arXiv preprint arXiv:1804.04338* (2018).
- [181] Christoph Baur, Shadi Albarqouni, and Nassir Navab. “Generating highly realistic images of skin lesions with GANs”. In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 260–267.
- [182] Mehdi Mirza and Simon Osindero. “Conditional generative adversarial nets”. In: *arXiv preprint arXiv:1411.1784* (2014).
- [183] Tero Karras et al. “Progressive growing of gans for improved quality, stability, and variation”. In: *arXiv preprint arXiv:1710.10196* (2017).
- [184] Baiying Lei et al. “Skin lesion segmentation via generative adversarial networks with dual discriminators”. In: *Medical Image Analysis* 64 (2020), p. 101716.
- [185] Devansh Bisla et al. “Towards automated melanoma detection with deep learning: Data purification and augmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 1–10.
- [186] Alceu Bissoto et al. “Skin lesion synthesis with generative adversarial networks”. In: *OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis*. Springer, 2018, pp. 294–302.

- [187] Federico Pollastri et al. “Augmenting data with GANs to segment melanoma skin lesions”. In: *Multimedia Tools and Applications* 79.21 (2020), pp. 15575–15592.
- [188] Sondre Fossen-Romsaas, Adrian Storm-Johannessen, and Alexander S Lundervold. “Synthesizing skin lesion images using CycleGANs—a case study”. In: *Norsk IKT-konferanse for forskning og utdanning*. 1. 2020.
- [189] Alceu Bissoto, Eduardo Valle, and Sandra Avila. “GAN-Based Data Augmentation and Anonymization for Skin-Lesion Analysis: A Critical Review”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 1847–1856.
- [190] Tero Karras, Samuli Laine, and Timo Aila. “A style-based generator architecture for generative adversarial networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4401–4410.
- [191] Ting-Chun Wang et al. “High-resolution image synthesis and semantic manipulation with conditional gans”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 8798–8807.
- [192] Taesung Park et al. “Semantic image synthesis with spatially-adaptive normalization”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2337–2346.
- [193] Lucy Chai, Jonas Wulff, and Phillip Isola. “Using latent space regression to analyze and leverage compositionality in GANs”. In: *arXiv preprint arXiv:2103.10426* (2021).
- [194] Tero Karras et al. “Training generative adversarial networks with limited data”. In: *arXiv preprint arXiv:2006.06676* (2020).
- [195] Tsendsuren Munkhdalai and Hong Yu. “Meta networks”. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 2554–2563.

- [196] Adam Santoro et al. “Meta-learning with memory-augmented neural networks”. In: *International conference on machine learning*. PMLR. 2016, pp. 1842–1850.
- [197] Nikhil Mishra et al. “A simple neural attentive meta-learner”. In: *arXiv preprint arXiv:1707.03141* (2017).
- [198] Oriol Vinyals et al. “Matching networks for one shot learning”. In: *Advances in neural information processing systems* 29 (2016), pp. 3630–3638.
- [199] Flood Sung et al. “Learning to compare: Relation network for few-shot learning”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 1199–1208.
- [200] Victor Garcia and Joan Bruna. “Few-shot learning with graph neural networks”. In: *arXiv preprint arXiv:1711.04043* (2017).
- [201] Jake Snell, Kevin Swersky, and Richard S Zemel. “Prototypical networks for few-shot learning”. In: *arXiv preprint arXiv:1703.05175* (2017).
- [202] Alex Nichol, Joshua Achiam, and John Schulman. “On first-order meta-learning algorithms”. In: *arXiv preprint arXiv:1803.02999* (2018).
- [203] Jaehong Kim et al. “Auto-meta: Automated gradient based meta learner search”. In: *arXiv preprint arXiv:1806.06927* (2018).
- [204] Edward Grefenstette et al. “Generalized inner loop meta-learning”. In: *arXiv preprint arXiv:1910.01727* (2019).
- [205] Sungyong Baik et al. “Meta-learning with adaptive hyperparameters”. In: *arXiv preprint arXiv:2011.00209* (2020).
- [206] Amy Zhao et al. “Data augmentation using learned transforms for one-shot medical image segmentation”. In: *CoRR* abs/1902.09383 (2019). arXiv: 1902.09383. URL: <http://arxiv.org/abs/1902.09383>.
- [207] Yunhui Guo et al. “A broader study of cross-domain few-shot learning”. In: *European Conference on Computer Vision*. Springer. 2020, pp. 124–141.

- [208] Massimiliano Patacchiola et al. “Bayesian Meta-Learning for the Few-Shot Setting via Deep Kernels”. In: *arXiv preprint arXiv:1910.05199* (2019).
- [209] John Cai, Bill Cai, and Sheng Mei Shen. “SB-MTL: Score-based Meta Transfer-Learning for Cross-Domain Few-Shot Learning”. In: *arXiv preprint arXiv:2012.01784* (2020).
- [210] Qianru Sun et al. “Meta-transfer learning for few-shot learning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 403–412.
- [211] Spyros Gidaris et al. “Boosting few-shot visual learning with self-supervision”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 8059–8068.
- [212] Timothy Hospedales et al. “Meta-learning in neural networks: A survey”. In: *arXiv preprint arXiv:2004.05439* (2020).
- [213] Yaqing Wang et al. “Generalizing from a few examples: A survey on few-shot learning”. In: *ACM Computing Surveys (CSUR)* 53.3 (2020), pp. 1–34.
- [214] Xiaomeng Li et al. “Difficulty-aware meta-learning for rare disease diagnosis”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 357–366.
- [215] Kushagra Mahajan, Monika Sharma, and Lovekesh Vig. “Meta-DermDiagnosis: few-shot skin disease identification using meta-learning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, pp. 730–731.
- [216] Delong Zhang, Mengqun Jin, and Peng Cao. “ST-MetaDiagnosis: Meta learning with Spatial Transform for rare skin disease Diagnosis”. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2020, pp. 2153–2160.

- [217] Dianwen Ng et al. “Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets”. In: *Quantitative Imaging in Medicine and Surgery* 11.2 (2021), p. 852.
- [218] Nicola Rieke et al. “The future of digital health with federated learning”. In: *NPJ digital medicine* 3.1 (2020), p. 119.
- [219] Bless Lord Y. Agbley et al. “Multimodal Melanoma Detection with Federated Learning”. In: *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. 2021, pp. 238–244. DOI: 10.1109/ICCWAMTIP53232.2021.9674116.
- [220] Md. Nazmul Hossen et al. “Federated Machine Learning for Detection of Skin Diseases and Enhancement of Internet of Medical Things (IoMT) Security”. In: *IEEE Journal of Biomedical and Health Informatics* 27.2 (2023), pp. 835–841. DOI: 10.1109/JBHI.2022.3149288.
- [221] Kyungsu Lee et al. “Fine-Tuning Network in Federated Learning for Personalized Skin Diagnosis”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 378–388.
- [222] Yawen Wu et al. “Federated Contrastive Learning for Dermatological Disease Diagnosis via On-device Learning (Invited Paper)”. In: *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*. 2021, pp. 1–7. DOI: 10.1109/ICCAD51958.2021.9643454.
- [223] Muhammad Mateen Yaqoob et al. “Federated Machine Learning for Skin Lesion Diagnosis: An Asynchronous and Weighted Approach”. In: *Diagnostics* 13.11 (2023), p. 1964.
- [224] Yanhang Shi, Xue Li, and Siguang Chen. “Skin Lesion Intelligent Diagnosis in Edge Computing Networks: A Federated Contrastive Learning Approach”. In: *ACM Transactions on Intelligent Systems and Technology* (2023).

- [225] Cho-I Moon et al. “Federated Learning for Masked Psoriasis Severity Classification”. In: *2022 IEEE Sensors. 2022*, pp. 1–4. DOI: 10.1109/SENSORS52175.2022.9967333.
- [226] Mohammed Adnan et al. “Federated learning and differential privacy for medical image analysis”. In: *Scientific reports* 12.1 (2022), p. 1953.
- [227] Micah J Sheller et al. “Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data”. In: *Scientific reports* 10.1 (2020), p. 12598.
- [228] Lawrence Sirovich and Michael Kirby. “Low-dimensional procedure for the characterization of human faces”. In: *Josa a* 4.3 (1987), pp. 519–524.
- [229] M Turk. “Pentland. Eigenfaces for recognition”. In: *K. Cogn. Neurosci* 4 (1991), pp. 72–86.
- [230] Ulf Grenander and Michael I Miller. *Pattern theory: from representation to inference*. OUP Oxford, 2006.
- [231] Thomas Gerig. “Gaussian Process Morphable Models for Spatially-Varying Multi-Scale Registration”. PhD thesis. University_of_Basel, 2021.
- [232] Pascal Paysan et al. “A 3D face model for pose and illumination invariant face recognition”. In: *2009 sixth IEEE international conference on advanced video and signal based surveillance*. Ieee. 2009, pp. 296–301.
- [233] Hanlin Zhang et al. “Rosacea treatment: review and update”. In: *Dermatology and Therapy* (2020), pp. 1–12.
- [234] Husein Husein-ElAhmed and Martin Steinhoff. “Laser and light-based therapies in the management of rosacea: an updated systematic review”. In: *Lasers in Medical Science* (2021), pp. 1–10.
- [235] Thomas Gerig et al. “Morphable face models-an open framework”. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE. 2018, pp. 75–82.

- [236] Kyle Genova et al. “Unsupervised training for 3d morphable model regression”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8377–8386.
- [237] Shiyang Cheng et al. “Meshgan: Non-linear 3d morphable models of faces”. In: *arXiv preprint arXiv:1903.10384* (2019).
- [238] Stylianos Moschoglou et al. “3dfacegan: Adversarial nets for 3d face representation, generation, and translation”. In: *International Journal of Computer Vision* 128.10 (2020), pp. 2534–2551.
- [239] Baris Gecer et al. “Fast-GANFIT: Generative Adversarial Network for High Fidelity 3D Face Reconstruction”. In: *arXiv preprint arXiv:2105.07474* (2021).
- [240] Ayush Jain et al. “Development and Assessment of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Tele dermatology Practices”. In: *JAMA network open* 4.4 (2021), e217249–e217249.
- [241] Maryam M Najafabadi et al. “Deep learning applications and challenges in big data analytics”. In: *Journal of big data* 2.1 (2015), pp. 1–21.
- [242] Neil Savage. “Synthetic data could be better than real data.” In: *Nature* (2023).
- [243] Martin Heusel et al. “Gans trained by a two time-scale update rule converge to a local nash equilibrium”. In: *Advances in neural information processing systems* 30 (2017).
- [244] Fisher Yu et al. “Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop”. In: *arXiv preprint arXiv:1506.03365* (2015).
- [245] Tuomas Kynkäänniemi et al. “Improved precision and recall metric for assessing generative models”. In: *Advances in Neural Information Processing Systems* 32 (2019).

- [246] Andrew Brock, Jeff Donahue, and Karen Simonyan. “Large scale GAN training for high fidelity natural image synthesis”. In: *arXiv preprint arXiv:1809.11096* (2018).
- [247] Mario Lucic et al. “Are gans created equal? a large-scale study”. In: *Advances in neural information processing systems* 31 (2018).
- [248] Tim Salimans et al. “Improved techniques for training gans”. In: *Advances in neural information processing systems* 29 (2016).
- [249] Tero Karras et al. *MetFaces Dataset*. URL: <https://metmuseum.github.io/>.
- [250] Alper Aksac et al. “BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis”. In: *BMC research notes* 12.1 (2019), pp. 1–3.
- [251] Alex Krizhevsky, Geoffrey Hinton, et al. “Learning multiple layers of features from tiny images”. In: *University of Toronto* (2009).
- [252] Ian Goodfellow et al. “Generative adversarial networks”. In: *Communications of the ACM* 63.11 (2020), pp. 139–144.
- [253] Jie Gui et al. “A review on generative adversarial networks: Algorithms, theory, and applications”. In: *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [254] Zhengli Zhao et al. “Image augmentations for gan training”. In: *arXiv preprint arXiv:2006.02595* (2020).
- [255] Zhengli Zhao et al. “Improved consistency regularization for gans”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 12. 2021, pp. 11033–11041.
- [256] Ashish Bora, Eric Price, and Alexandros G. Dimakis. “AmbientGAN: Generative models from lossy measurements”. In: *International Conference on Learning Representations*. 2018, pp. 1–22. URL: <https://openreview.net/forum?id=Hy7fDog0b>.

- [257] Kevin Roth et al. “Stabilizing training of generative adversarial networks through regularization”. In: *Advances in neural information processing systems* 30 (2017).
- [258] William Fedus et al. “Many paths to equilibrium: GANs do not need to decrease a divergence at every step”. In: *arXiv preprint arXiv:1710.08446* (2017).
- [259] Charles A Holt and Alvin E Roth. “The Nash equilibrium: A perspective”. In: *Proceedings of the National Academy of Sciences* 101.12 (2004), pp. 3999–4002.
- [260] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. “Which training methods for GANs do actually converge?” In: *International conference on machine learning*. PMLR. 2018, pp. 3481–3490.
- [261] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. “The numerics of gans”. In: *Advances in neural information processing systems* 30 (2017).
- [262] Anwesha Mohanty et al. “Skin disease analysis with limited data in particular Rosacea: a review and recommended framework”. In: *IEEE Access* 10 (2022), pp. 39045–39068.
- [263] Frank Powell. *Powell Lab*. URL: <https://www.ucd.ie/charles/research/researchgroups/thepowelllab/>.
- [264] Charles Institute of Dermatology. *Charles Institute of Dermatology, University College Dublin*. URL: <https://www.ucd.ie/charles/>.
- [265] Mikołaj Bińkowski et al. “Demystifying mmd gans”. In: *arXiv preprint arXiv:1801.01401* (2018).
- [266] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. “Freeze the discriminator: a simple baseline for fine-tuning gans”. In: *arXiv preprint arXiv:2002.10964* (2020).
- [267] Ali Borji. “Pros and cons of GAN evaluation measures: New developments”. In: *Computer Vision and Image Understanding* 215 (2022), p. 103329.

- [268] Han Zhao et al. “Learning neural networks with adaptive regularization”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [269] Anna Escalé-Besa et al. “Exploring the potential of artificial intelligence in improving skin lesion diagnosis in primary care”. In: *Scientific Reports* 13.1 (2023), p. 4293.
- [270] Andre Esteva et al. “A guide to deep learning in healthcare”. In: *Nature medicine* 25.1 (2019), pp. 24–29.
- [271] Jiajun Wu et al. “Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling”. In: *Advances in neural information processing systems* 29 (2016).
- [272] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *Communications of the ACM* 65.1 (2021), pp. 99–106.
- [273] Feng Fu. “Design and Analysis of Complex Structures”. In: *Design and Analysis of Tall and Complex Structures; Butterworth-Heinemann: Oxford, UK* (2018), pp. 177–211.
- [274] Tobias Heimann and Hans-Peter Meinzer. “Statistical shape models for 3D medical image segmentation: a review”. In: *Medical image analysis* 13.4 (2009), pp. 543–563.
- [275] Nicholas J Mankovich et al. “Surgical planning using three-dimensional imaging and computer modeling”. In: *Otolaryngologic Clinics of North America* 27.5 (1994), pp. 875–889.
- [276] Anil Murat Ozturk et al. “Multidisciplinary assessment of planning and resection of complex bone tumor using patient-specific 3D model”. In: *Indian Journal of Surgical Oncology* 10 (2019), pp. 115–124.
- [277] Ruvi Chauhan, Scott N Loewenstein, and Aladdin H Hassanein. “Rhino-phyma: prevalence, severity, impact and management”. In: *Clinical, cosmetic and investigational dermatology* (2020), pp. 537–551.

- [278] ROKEA A el-AZHARY, RANDALL K ROENIGK, and TOM D WANG. “Spectrum of results after treatment of rhinophyma with the carbon dioxide laser”. In: *Mayo Clinic Proceedings*. Vol. 66. Elsevier. 1991, pp. 899–905.
- [279] T Wetzig et al. “New rhinophyma severity index and mid-term results following shave excision of rhinophyma”. In: *Dermatology* 227.1 (2013), pp. 31–36.
- [280] Krisztina Somogyvári et al. “Radiosurgical excision of rhinophyma”. In: *Dermatologic surgery* 37.5 (2011), pp. 684–687.
- [281] Abid Haleem and Mohd Javaid. “3D scanning applications in medical field: a literature-based review”. In: *Clinical Epidemiology and Global Health* 7.2 (2019), pp. 199–210.
- [282] Maria HJ Hollander et al. “Reproducibility of 3D scanning in the periorbital region”. In: *Scientific Reports* 11.1 (2021), p. 3671.
- [283] Dermatoweb.net. *Dermato web*. Available at <http://dermatoweb.net>.
- [284] Blender Online Community. *Blender - A 3D Modelling and Rendering Package*. Blender Foundation. Stichting Blender Foundation, Amsterdam, 2018. URL: <http://www.blender.org>.
- [285] National Electrical Manufacturers Association et al. “Digital imaging and communication in medicine (DICOM)”. In: *NEMA PS 3 Supplement 23 Structured Reporting* (1997).
- [286] 3D Slicer. *3D Slicer image computing platform*. Available at <https://www.slicer.org/>.
- [287] Brain 3D. *Brain 3D*. Available at <https://www.brain3d.co/>.
- [288] Vahid Kazemi and Josephine Sullivan. “One Millisecond Face Alignment with an Ensemble of Regression Trees”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 1867–1874.

- [289] dlib. *dlib*. Available at http://dlib.net/face_landmark_detection.py.html.
- [290] Mark Sandler et al. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520.
- [291] Andrew Howard et al. “Searching for mobilenetv3”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 1314–1324.
- [292] Laurent Sifre and Stéphane Mallat. “Rigid-motion scattering for texture classification”. In: *arXiv preprint arXiv:1403.1687* (2014).
- [293] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. “Distilling the knowledge in a neural network”. In: *arXiv preprint arXiv:1503.02531* (2015).
- [294] Forrest N Iandola et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5 MB model size”. In: *arXiv preprint arXiv:1602.07360* (2016).
- [295] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.
- [296] OpenCV. *Haar cascade classifier*. Available at https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html.
- [297] Mohanty Anwesha et al. *3D-rhi-synth-2000- Synthetic Rhinophyma Visual Dataset*. Version v1. Aug. 2023. DOI: 10.5281/zenodo.8228258. URL: <https://doi.org/10.5281/zenodo.8228258>.
- [298] Amir Al-Dabagh et al. “Rosacea in skin of color: not a rare diagnosis”. In: *Dermatology Online Journal* 20.10 (2014).
- [299] Rashmi Sarkar, Indrashis Podder, and Soumya Jagadeesan. “Rosacea in skin of color: A comprehensive review”. In: *Indian Journal of Dermatology, Venereology and Leprology* 86 (2020), p. 611.

- [300] T Maruthappu and M Taylor. “Acne and rosacea in skin of colour”. In: *Clinical and Experimental Dermatology* 47.2 (2022), pp. 259–263.
- [301] Andrew F Alexis et al. “Global epidemiology and clinical spectrum of rosacea, highlighting skin of color: review and clinical practice experience”. In: *Journal of the American Academy of Dermatology* 80.6 (2019), pp. 1722–1729.
- [302] Yongcheng Jing et al. “Neural style transfer: A review”. In: *IEEE transactions on visualization and computer graphics* 26.11 (2019), pp. 3365–3385.
- [303] Zhenliang He et al. “Attgan: Facial attribute editing by only changing what you want”. In: *IEEE transactions on image processing* 28.11 (2019), pp. 5464–5478.
- [304] Ming Liu et al. “Stgan: A unified selective transfer network for arbitrary image attribute editing”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 3673–3682.
- [305] Paul Upchurch et al. “Deep feature interpolation for image content changes”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7064–7073.
- [306] Pang Yun et al. “A Review of Face Attribute Editing Based on Deep Learning”. In: *Proceedings of the 2023 15th International Conference on Machine Learning and Computing*. 2023, pp. 127–133.
- [307] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [308] Alec Radford et al. “Learning transferable visual models from natural language supervision”. In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763.
- [309] Nan Luo et al. “Artificial intelligence-assisted dermatology diagnosis: from unimodal to multimodal”. In: *Computers in Biology and Medicine* (2023), p. 107413.