International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 6
ISSN: 2088-5334

Facial Skin Type Analysis Using Few-shot Learning with Prototypical Networks

Quan Fong Yeo [a], Shih Yin Ooi [a,*], Ying Han Pang [a], Ying Huey Gan [a]

[a] *Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka, 75450, Melaka, Malaysia*

*Corresponding author: *syooi@mmu.edu.my*

Abstract*—Facial skin type analysis is a critical task in several fields, including dermatology, cosmetics, and biometrics, and has been the subject of significant research in recent years. Traditional facial skin type analysis approaches rely on large, labeled datasets, which can be time-consuming and costly to collect. This study proposes a novel few-shot learning (FSL) approach for facial skin type analysis that can accurately classify skin types with limited labeled data. A diverse dataset of facial images with varying skin tones and conditions was curated. The proposed approach leverages pre-trained deep neural networks and an FSL algorithm based on prototypical networks (PNs) and matching networks (MNs) to address the challenge of limited labeled data. Importantly, this study has significant implications for improving access to dermatological care, especially in underserved populations, as many individuals are unaware of their skin type, which can lead to ineffective or even harmful skincare practices. Our approach can help individuals quickly determine their skin type and develop a personalized skincare routine based on their unique skin characteristics. The results of our experiments demonstrate the effectiveness of the proposed approach. PNs achieved the highest accuracy in the 2-way, 10-shot, 15-query scenario with an accuracy of 95.78 ± 2.79%, while MNs achieved the highest accuracy of 90.33 ± 4.10% in the 2-way, 5-shot, 10-query scenario. In conclusion, this study highlights the potential of FSL and deep neural networks to overcome the limitations of traditional approaches to facial skin analysis, offering a promising avenue for future research in this field.

***Keywords*—Skin type analysis; skincare products recommendation; deep learning; few-shot learning; convolutional neural network.**

Manuscript received 11 Jan. 2023; revised 24 Jul. 2023; accepted 9 Sep. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Analyzing facial skin types is paramount in various domains, including dermatology, cosmetics, and biometrics. Extensive research efforts have been dedicated to understanding skin characteristics and developing effective skincare practices through facial skin type analysis. However, traditional approaches to this analysis rely heavily on large, labeled datasets, which pose challenges in terms of time and resources for collection [1]. Consequently, alternative methodologies are needed to classify skin types with limited labeled data accurately.

In addition, the specific context of Malaysian facial skin types further compounds the challenges in this field [2], [3], [4]. There is a notable absence of a labeled facial dataset tailored to Malaysian skin types [5], [6]. Moreover, existing applications that aim to determine skin type often lack a specific focus on Malaysian individuals. Many popular applications, such as L'Oréal Paris Skin Genius, Olay Skin Advisor, and others that belong to well-known brands, primarily cater to a broader audience and do not adequately represent the diverse range of Malaysian skin types. This significant gap in dedicated resources for Malaysian facial skin types highlights the need for our collection of a comprehensive Malaysian facial dataset.

Therefore, this paper proposes a novel few-shot learning (FSL) approach for facial skin type analysis with prototypical networks (PNs) and matching networks (MNs) to overcome these limitations [7], [8]. By leveraging the power of these networks, we aim to address the challenges associated with traditional methods of classifying facial skin types [9]. Our approach enables accurate classification of skin types even with a limited amount of labeled data, thus reducing reliance on extensive labeled datasets. This advancement solves the scarcity of labeled facial datasets and opens new possibilities for accurate facial skin type analysis in the Malaysian context.

The research is motivated by its critical implications for improving access to dermatological care, particularly among underserved populations [10], [11]. It is common for individuals to be unaware of their skin type, leading to ineffective or potentially harmful skincare practices [12],

[13]. By addressing the challenges of limited labeled data and the lack of dedicated applications for Malaysian facial skin types, our proposed approach offers individuals an easy and reliable method to determine their skin type. This, in turn, empowers them to develop personalized skincare routines based on their unique skin characteristics, contributing to improved skincare practices and overall well-being among the Malaysian population [14], [15].

In Section 2, we present an extensive literature review, discussing previous research on facial skin type analysis, highlighting the limitations of traditional approaches, and exploring the potential of FSL in this context. Section 3 outlines the data collection and pre-processing procedures, including creating a diverse dataset of facial images and the necessary annotations. Section 4 presents the methodology, providing a detailed explanation of the proposed approach utilizing PNs and FSL. Section 5 presents the experimental setup and results. In contrast, Section 6 concludes the paper by summarizing the findings, suggesting future research directions in dermatology cosmetics, and suggesting avenues for future research.

A. Justification of Study

This literature review aims to investigate the application of FSL methods, particularly utilizing PNs, in the analysis and classification of facial skin types. While existing research predominantly focuses on skin disease classification [16]-[24], limited studies specifically address facial skin type classification. By examining relevant studies on facial skin type analysis and adapting their methodologies, we aim to bridge this research gap and develop accurate and efficient approaches for facial skin type analysis. The review will compare traditional machine learning algorithms, deep learning architectures, and state-of-the-art skin analysis apps, ultimately providing insights and establishing a foundation for our proposed approach utilizing FSL and PNs in facial skin type classification [25], [26].

1) Facial Skin Typing System: Skin typing systems are crucial in classifying and categorizing different skin types based on various parameters, providing a framework for understanding and addressing individual skin characteristics in personalized skincare approaches. Several skin typing systems have been developed and utilized in dermatology and skincare [1].

One widely recognized skin typing system is the Baumann Skin Type System (BSTS), which incorporates four main parameters: dry or oily, sensitive or resistant, pigmented or non-pigmented, and wrinkle-prone or tight. Combining these parameters, the BSTS creates 16 distinct skin types, offering a comprehensive classification approach [27]. The BSTS employs a validated questionnaire, the Baumann Skin Type Indicator (BSTI), to assess and assign individuals to their skin types. This versatile system applies to all ethnicities, ages, and genders, comprehensively understanding diverse skin characteristics [28]–[33].

Another commonly used skin typing system is the Fitzpatrick scale, which primarily focuses on categorizing skin based on its response to sun exposure and its tendency to tan or burn. The Fitzpatrick scale classifies skin into six types ranging from very fair (Type I) to very dark (Type VI). This system is particularly relevant in determining the risk of sun damage, guiding appropriate sun protection measures, and predicting the likelihood of certain skin conditions, such as skin cancer [34]–[37].

The Roberts Skin Type Classification System is another comprehensive approach that evaluates four elements: phototype, hyperpigmentation, photoaging, and scarring, to classify an individual's skin type and predict their response to insult and inflammation. This system combines quantitative and qualitative assessments, including ancestral and clinical history, visual examination, test site reactions, and physical examination. It provides valuable information for treatment planning, managing patient expectations, and optimizing outcomes, ultimately improving physician-patient communication, patient compliance, and preventive measures [38].

During the early 1900s, Helena Rubinstein introduced a primary system for classifying skin types, which has since remained a traditional and widely utilized approach in the skincare industry. This system establishes four fundamental skin types: normal, dry, oily, and sensitive. While these classifications have been valuable in providing a general understanding of skin characteristics, scholars have raised concerns about their limited capacity to encompass intricate descriptions such as pigmentation irregularities or the presence of wrinkles [27].

In addition to these established systems, various other skin typing approaches exist, emphasizing specific aspects of the skin, such as sebum production, skin barrier function, or acne susceptibility. These systems aim to understand individual skin characteristics better and address specific skincare concerns [1]. However, this study employs a simplified skin typing system to evaluate the viability of FSL techniques for classifying skin types. Thus, the traditional skin typing system by Helena Rubinstein is chosen, and the efficacy of FSL methodologies in achieving accurate skin type classification will be assessed.

2) Few-shot Learning (FSL): FSL employs the N-way K-shot classification approach, which aims to differentiate between N classes using K examples. Specifically, the N- way K-shot image classification task involves a support set containing K-labelled images. Additionally, a query set comprising Q query images is provided. The goal is to accurately classify the query images into the N classes based on the information provided by the $N \times K$ images in the support set. In the context of FSL, K is typically a small value, often less than 10. When K equals 1, it is commonly called one-shot classification [39]–[42].

Fig. 1 provides an example of a few-shot classification task. In this case, the support set includes $K = 2$ instances for each of the $N = 3$ classes: duck, penguin, and chicken. The objective is to assign labels to the Q = 4 birds in the query set accurately, categorizing them as duck, penguin, or chicken. While humans can effortlessly accomplish this task without familiarity with these specific bird species, solving it using artificial intelligence requires meta-learning techniques to achieve successful classification [43].

Fig. 1 Few-shot classification task example.

FSL can be considered a form of meta-learning or learning to learn, as it involves acquiring experience from other related problems [44]–[50]. Recent FSL advancements have been achieved primarily through applying meta-learning techniques, especially metric-based meta-learning [51]–[54]. Therefore, this paper is dedicated to exploring and discussing metric-based meta-learning, as it aligns with the chosen approach for this study.

Metric learning involves learning a distance function between data points, such as images. The core idea is to extract embeddings from all images in both the support and query sets with a Convolutional Neural Network (CNN) and assign the label corresponding to the image with the shortest distance, similar to k-nearest neighbors' algorithms (k-NN). At the end of each episode, the CNN parameters are updated through backpropagation of the loss computed from the classification error on the query set, typically using a cross-entropy loss.

Metric learning is prevalent in few-shot image classification due to its proven effectiveness and abundant opportunities for innovation and improvement in feature extraction and comparison techniques [55]–[59]. In the upcoming section, we will dive into a few existing solutions.

Matching Networks (MNs) are the first metric learning algorithm using meta-learning, MNs presents a meta-learning-based metric learning algorithm for few-shot image classification. It adopts distinct feature extraction procedures for the support set images and query images. The query embeddings are then compared to each support set image using cosine similarity and classified through a SoftMax operation. Notably, the authors propose using Long Short-term Memory Networks (LSTM) to enable comprehensive interaction among all images during the feature extraction process, referred to as Full Context Embedding [8]. This approach demonstrates improved performance compared to a simple CNN-based approach but at the cost of increased computational requirements and GPU resources [8], [60], [61]. [8, Fig. 2] illustrates the architecture of MNs, where separate feature extractors are employed for the support set images on the left and query images at the bottom. The embedding of the query image is compared to each image in the support set using cosine similarity. Then, a SoftMax classification is performed to assign a class label to the query based on the computed similarities.

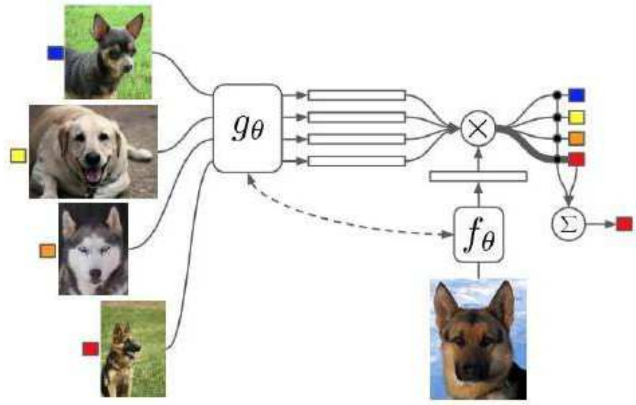


Fig. 2 MNs architecture. Adapted from [8]

Next, recent studies of Prototypical Networks (PNs) have shifted away from the traditional approach of comparing query images with every image in the support set. An example of such a departure is the introduction of PNs by Snell et al. [7], [62]–[66]. This metric learning algorithm computes prototypes for each class by averaging the embeddings of all images within the class, allowing for various ways of computing these embeddings as long as the function is differentiable. Subsequently, queries are classified based on their Euclidean distance to the prototypes. PNs have demonstrated state-of-the-art performance in few-shot image classification tasks despite their simplicity. While more complex metric-learning architectures, such as neural networks representing the distance function, have been developed with slight accuracy improvements, prototypes remain highly valuable in metric-learning algorithms for few-shot image classification [67]–[73]. Fig. 3 demonstrates that the few-shot prototypes are computed as the means of embedded support examples for each class.

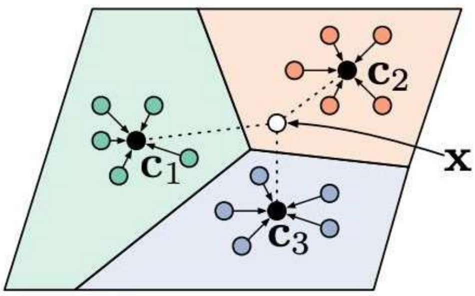


Fig. 3 PNs in the FSL scenario. Adapted from [7].

The use of few-shot learning methods in skin-related studies has shown promise and can be applied to the classification of facial skin types, which is one of the contributions of this paper. Although there is a lack of specific research on determining facial skin types using FSL, the methods employed in skin disease image classification can offer valuable insights and potentially be adapted for this purpose.

For instance, Prabhu et al. [74] utilized Prototypical Clustering Networks (PCN) to classify the age of dermatological images. Mahajan et al. [75] employed meta-learning techniques and G-convolutions to successfully identify skin diseases, surpassing the performance of previous methods. Additionally, Liu et al. [76] achieved remarkable accuracy in categorizing skin lesion images using FSL. These studies demonstrate the effectiveness and potential of few-shot learning in skin-related classification tasks. Therefore, drawing inspiration from these approaches, this paper aims to leverage FSL techniques for classifying facial skin types.

II. Materials and Method

This part of the research paper introduces how a machine-learning classifier can analyze skin texture. It provides detailed insights into the designed approach in the following sections.

A. Dataset Collection

For this research, the data collection utilizes a skin typing system by Helena Rubinstein: normal, dry, oily, and sensitive. This method is suitable for testing the original skin types found in Malaysia due to the limited availability of existing data. This research aims to contribute to creating a comprehensive dataset that represents the diverse range of skin types prevalent in Malaysia and Asia. Our data collection aims to support the skincare industry in Malaysia and Asia, two of the most significant beauty and personal skincare markets globally [77]. Gathering a valuable dataset enables research, innovation, and collaboration in skincare, benefiting the industry and advancing global progress. Through the data collection from Malaysia and Asia, we contribute to enhancing skincare practices and boosting well-being in the region.

1) Questionnaires: The dataset collection process commences at Multimedia University (MMU) Melaka Campus, where participants were invited to participate in the study. The initial data collection phase occurred in a controlled environment, utilizing a booth set up at MMU. This environment allows precise control over image quality, lighting conditions, and participant positioning.

Within this controlled environment, participants were requested to complete the modified version of the BSTI questionnaire [28], consisting of 15 questions tailored to align with the tone and context suitable for university students. The questionnaire aims to efficiently capture essential information about participants' skin conditions while ensuring a precise classification into the skin types of normal, dry, oily, and sensitive without directly replicating the wording of the original version. The modified version of the questionnaire is presented in Table 4 in Appendix A of this paper.

Completing the modified BSTI questionnaire typically took participants approximately 3 to 5 minutes. This concise timeframe minimizes the burden on participants while allowing for efficient data collection. Following the questionnaire, participants' facial images were captured using a designated device, the Canon EOS Rebel T4i, an 18.0-megapixel digital single-lens reflex (DSLR) camera, to ensure consistent and standardized image quality. The device is positioned carefully to ensure accurate framing and optimal lighting during image capture. This approach aims to minimize variations that could affect the quality and comparability of facial images. Throughout the process, participants were provided with supervision and assistance to ensure that their questions and concerns were addressed promptly, creating a supportive and comfortable environment and enhancing the quality and accuracy of the collected data.

In parallel, data collection was also conducted in an uncontrolled environment using an online platform. A dedicated Google site (https://sites.google.com/view/msftd-survey/home) has been created to inform participants about the research project and provide instructions for uploading their photos. Participants followed specific guidelines when capturing their photos, including removing makeup, removing glasses and tying back hair, maintaining a neutral expression, and ensuring normal lighting conditions. Additionally, participants in the uncontrolled environment must complete the modified BSTI questionnaire and upload their photos through the provided domain. At the same time, the uncontrolled environment represents real-world conditions with variations in lighting and environmental factors. This dual approach of controlled and uncontrolled environments helps capture a diverse range of facial images and mitigate the effects of domain shift.

By initiating the dataset collection process at MMU and employing both controlled and uncontrolled environments, we aim to gather a comprehensive dataset encompassing various skin types and conditions for further analysis and research. After the initial phase of dataset collection, a total of 218 participants were involved. Among them, 140 participants were in a controlled environment, while 78 were in an uncontrolled environment. The gender distribution consisted of 137 men and 81 women. All participants were aged between 18 and 24 years.

2) Calculation Process for Determining Participant's Skin Type: In this subsection, we outline the calculation process of classifying skin types based on the scores obtained from the questionnaire responses. The questionnaire is divided into four sections based on BSTS's four main parameters for determining skin type. Section 1 includes questions 1-7 assessing the dry or oily parameter. Section 2 comprises questions 8-11, focusing on the sensitive or resistant parameter. Section 3 addresses the pigmented or non-pigmented parameter and consists of questions 12-13. Finally, Section 4 examines the wrinkle-prone or tight parameter and includes questions 14-15.

After participants completed the questionnaire, the collected data was analyzed using a scoring system to assign individuals to different skin types. The survey assigned an equal weight to each question since they were all considered equally relevant in determining the participant's skin type. A scoring system was implemented, assigning each answer option a specific point value.

In Section 1 of the questionnaire, the scoring system assigned point values to each answer option as follows: "a" = 1 point, "b" = 2 points, "c" = 3 points, and "d" = 4 points. However, if a participant selected option "e" for any question, it indicated that they were unsure or did not provide a response. In such cases, those specific questions were omitted from the calculation. Adjustments were made to the score range to accommodate this omission as shown in Table 1 and ensure accurate determination of the participant's skin type.

TABLE I
CUT OFF POINTS FOR DETERMINING SKIN TYPE IN SECTION 1 OF THE QUESTIONNAIRE.

Option “e” Count	Dry Skin Range	Oily Skin Range
0	7-14	15-28
1	6-12	13-24
2	5-10	11-20
3	4-18	9-16
4	3-6	7-12

In Section 2, the scoring system assigned point values to each answer option as follows: “a” = 1 point, “b” = 2 points, “c” = 3 points, and “d” = 4 points. Participants scoring between 4 and 8 were categorized as having sensitive skin, while scores ranging from 9 to 16 indicated resistant skin. In Section 3, if the participant answered “yes” to question 12, it signified the presence of wrinkles and “no” to represent tight. Lastly, in Section 4, if the participant selected “none” for pigmented in question 13, it indicated the absence of pigmentation. By following this step-by-step calculation process for each parameter, the participant’s BSTI was determined.

3) Classification of Helena Rubinstein Skin Types Based on Modified BSTI Questionnaire: After classifying participants into their BSTI, the next step is to convert the BSTI into the Helena Rubinstein skin typing system: normal, dry, oily, and sensitive. The categorizations are as follows:

- Normal: The normal skin type was defined as ORNT [78]. Participants with ORNT skin type were considered to have a normal skin type in the context of this paper as they have normal sebum production and resistance to environmental factors.
- Dry: Participants classified as dry (D) indicators were categorized as having dry skin.
- Oily: Participants classified as oily (O) indicators were categorized as having oily skin.
- Sensitive: To simplify the analysis, participants classified as sensitive (S) indicators were categorized as having sensitive skin regardless of their specific dry or oily classification. This category included participants with skin types such as DSPT, DSNT, DSPW, DSNW, OSPT, OSNT, OSPW, and OSNW, as all these skin types had sensitive parameters.

The decision to exclude the two parameters: pigmented (P) or non-pigmented (N) and wrinkle-prone (W) or tight (T) parameters from specific categorization in this study was made to maintain the focus on the normal, dry, oily, and sensitive skin types outlined in the Helena Rubinstein system. Instead, skin types that exhibited characteristics under the these two parameters were considered part of dry or oily skin types. For example, in this paper, ORNW and ORNT were both classified as oily skin; ORPW and ORNW were both classified as oily skin in this paper, regardless of their P or N and W or T parameters.

This approach aimed to simplify the analysis and maintain the research objectives of focusing on normal, dry, oily, and sensitive skin types. It is important to note that skin type classifications may vary depending on the specific system or parameters considered. The categorizations made in this study were designed to align with the research objectives and provide clarity and practicality in analyzing participants’ skin characteristics. Additional parameters, such as wrinkled-prone or tight and pigmented or non-pigmented, could be explored in future studies for a more comprehensive understanding of skin types.

After analyzing the questionnaire responses and conducting calculations, the dataset consisted of 33 participants with normal skin (15%), 50 with dry skin (23%), 74 with oily skin (34%), and 61 with sensitive skin (28%). Additionally, the participants represented diverse ethnic backgrounds, including Malay, Chinese, and Indian. These demographic details provide valuable insights into the sample population and ensure a diverse representation of various skin types and ethnicities within the dataset. This diversity enhances the robustness and applicability of the dataset for future studies in the field.

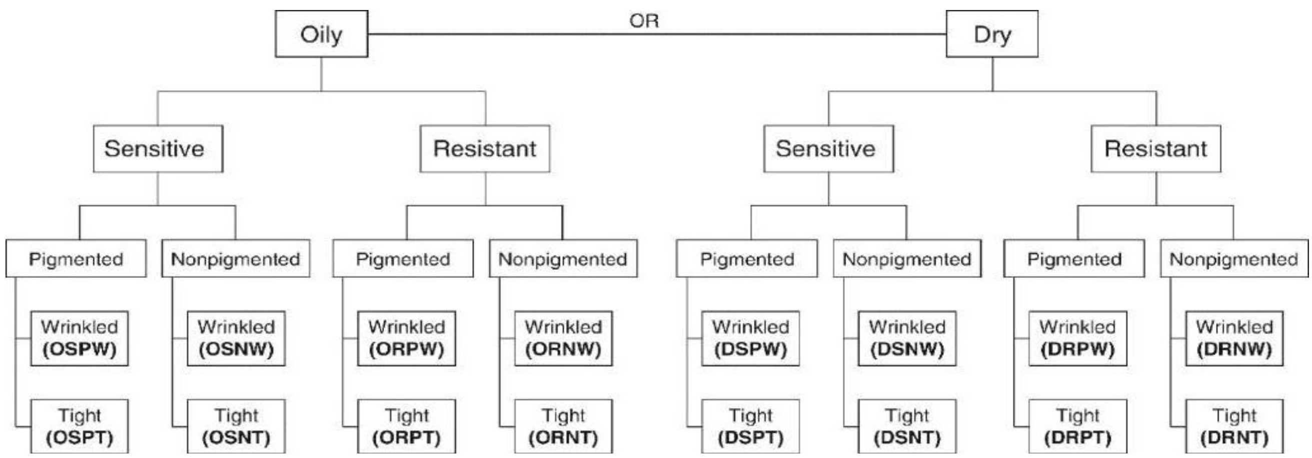


Fig. 4 The BSTI skin types. Adapted from [79].

Our skin type classification in the dataset was based on participant survey responses following established frameworks and literature reviews. While we acknowledge that self-reported skin type may have varying degrees of accuracy, we took necessary precautions and provided comprehensive information to enhance participants’ understanding. Although expert verification was not conducted, relying on participant responses is common in many studies involving skin type classification.

To enhance participants' understanding, we have provided comprehensive information on our website to help them grasp the study's purpose, process, and the relevant concepts associated with skin typing. Our intention is to ensure transparency and provide participants with the necessary knowledge to make informed responses.

It is important to note that self-reported skin types may have varying degrees of accuracy, and further research or expert assessment may be necessary for more precise categorization. However, we have strived to ensure the reliability and validity of our findings by conducting extensive research, reviewing relevant literature, and seeking guidance from reliable sources during the questionnaire design process.

B. Image Pre-processing

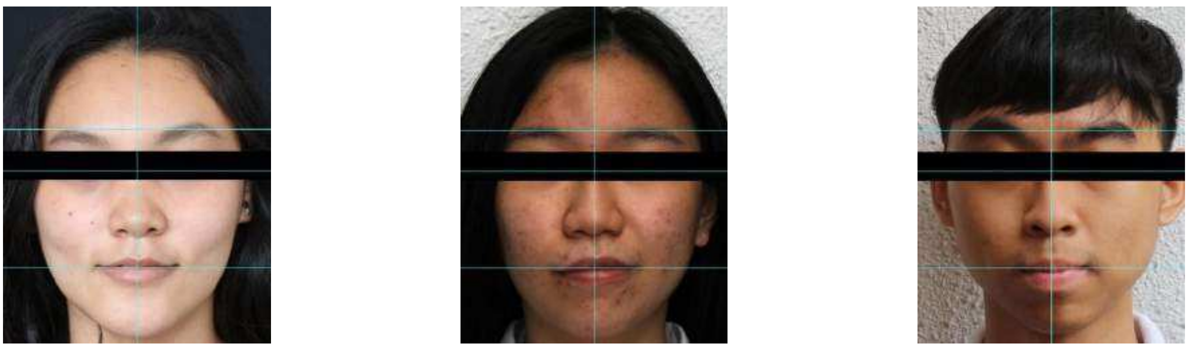
Various techniques are applied during the pre-processing step of the collected facial images to enhance their quality and ensure consistency across the dataset. First, intelligent filters are applied to retain the adjustment settings, allowing reproducibility of the pre-processing steps. This enables the retention of filter settings and facilitates further refinements if necessary. Additionally, normalization is performed to standardize the range of values in the images, minimizing variations in lighting conditions and color intensity.

An algorithm within Adobe Photoshop matches the colour tone, contrast, and lighting between images captured in controlled and uncontrolled environments. This algorithm analyses the characteristics of the controlled images and applies appropriate adjustments to the uncontrolled images, aligning them as closely as possible. This process helps create a uniform appearance and visual consistency across the dataset, reducing visual disparities caused by the domain shift. Furthermore, a spatial transformation technique is implemented to resize the images and align each individual's facial features. This step involves adjusting the size and positioning of the facial features to ensure consistency and accuracy. The images are standardized by aligning the facial features, enabling accurate analysis and comparison of various facial attributes across all domains.

After the image processing and segmentation, the dataset of 218 participants was divided into 1,962 facial part images. Among the participants, 33 had normal skin, resulting in 297 normal skin facial part images. There were 50 participants with dry skin, resulting in 450 dry skin facial part images. Additionally, 74 participants had oily skin, contributing 666 oily skin facial part images. Finally, there were 61 participants with sensitive skin, resulting in 549 sensitive skin facial part images.

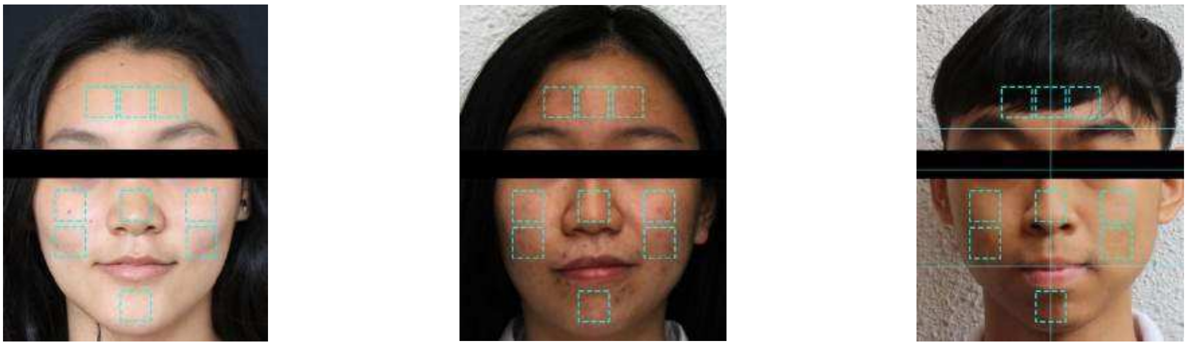
(a) (b) (c)

Fig. 5 Example images illustrating the impact of different environments on image quality and appearance. In (a) image captured in a controlled environment showing normal skin, (b) image captured in an uncontrolled environment showing sensitive skin, and (c) image captured in an uncontrolled environment showing dry skin. The differences in lighting, image quality, and environmental factors are apparent. All participants have consented to this project, and a black bar has been added to protect their anonymity.



(a) (b) (c)

Fig. 6 Examples of image pre-processing techniques applied to images (a), (b), and (c) to align facial features through spatial transformation, resizing, and adjusting color tone to match natural lighting conditions. These pre-processing steps ensure consistency in facial feature alignment, enhance visual consistency, and improve the overall image quality for subsequent analysis.



(a) (b) (c)

Fig. 7 Segmentation based on facial features. The segmented areas include the forehead, right cheek, left cheek, nose, and chin. A total of 9 images per participant are extracted, each representing a specific facial region.

(a) (b) (c)

Fig. 8 Example of images after segmentation and resizing. The images are resized to 84x84 pixels to facilitate training with limited computational resources. Each row represents a different facial region: the first row corresponds to the forehead, the second row to the right cheek, the third row to the left cheek, the fourth row to the nose, and the fifth row to the chin. Before training, a careful selection process is conducted to ensure image quality and suitability. For instance, in the image (c) of the forehead, the presence of hair obscuring the region leads to its exclusion from the dataset. Similarly, any other obscured images are also discarded during the selection process.

Following the selection process to exclude obscured images, the final dataset consisted of 259 normal skin facial part images, 408 dry facial skin part images, 592 oily skin facial part images, and 516 sensitive skin facial part images. A random selection of 300 images from each class was made to ensure a balanced dataset for few-shot learning. Data augmentation techniques [80] were then applied to the normal skin class to expand the dataset to 300 images while maintaining the class balance among the other classes.

By keeping the class balance, the dataset enables effective few-shot learning, where models can be trained with limited examples from each class. This approach promotes better generalization and performance when faced with new and unseen facial skin images. The initial version of the Malaysian Facial Skin Texture Dataset (MFSTD) has been completed. The dataset comprises 1,200 images, with 300 images per class representing normal, dry, oily, and sensitive skin types.

C. Implementation

In this section, we provide a detailed overview of the implementation process for our FSL approach. We cover data pre-processing, architecture selection, model training, and optimization techniques employed in our experiments.

1) Data Pre-processing: In the context of traditional few-shot learning benchmark datasets, a common practice involves splitting the data into predefined train, test, and validation sets with a disjoint distribution of classes [81]. However, when working with our dataset consisting of only four classes, it is not possible to maintain a disjoint distribution of classes across these sets. Consequently, we adopt a modified approach to ensure the evaluation of model generalization, which split our dataset into three sets: train, test, and validation, using a ratio of 6:2:2, respectively. This distribution ensured that 60% of the data was assigned to the training set, while 20% each was allocated to the test and validation sets. This division also allowed us a sufficiently large training set to train our models while ensuring a fair evaluation of the test and validation sets.

It is important to note that although the train, test, and validation sets shared the same four classes, each set contained distinct images. This ensured that the model was exposed to different samples during training, testing, and validation. While this approach deviates from the traditional few-shot learning setting, where disjoint class distributions are typically used, we found support for our methodology, stating that disjoint classes are not a technical requirement, thereby validating our approach [43].

One concern with the original disjoint approach, particularly in our case, is that assigning only one class to the validation set will expose the model to the same class it needs to predict, resulting in potentially inflated accuracy scores. This is because the FSL scenario involves training the model on a small support set and evaluating its performance on a query set consisting of samples from the same class. Thus, the model is expected to achieve high accuracy on the validation set due to its familiarity with the specific class. While this modified approach may yield higher accuracy on the validation set due to the shared classes, it still provides valuable insights into the model's ability to learn and generalize within those classes.

For all the datasets used in our experiments, we uniformly resized the images to 84x84 pixels, in line with the specifications of the mini-ImageNet dataset. This standard image size allows for consistent and comparable analysis across different datasets and ensures that the images suit our specific model and computational resources.

2) Architecture: In all our experiments, we use the custom Residual Neural Network 12 (ResNet12) architecture for representing the embedding function. This custom ResNet12 architecture consists of four blocks with widths [64, 160, 320, 640]. This specific architecture was chosen based on its frequent adoption by recent FSL methods, initially introduced by [47]. Unlike many other methods, our experiments did not incorporate the DropBlock regulariser. This decision was based on the findings of Ghiasi et al. [82], who reported that the inclusion of DropBlock did not significantly contribute to the performance of their models.

3) Training Model: To ensure a fair comparison between PNs and MNs, we construct the episodes by aligning the number of shots during training and testing [8], [81]. In our evaluation, episodes consist of 2-, 3-, or 4-way classification and 1-, 5-, or 10-shot learning, with 5-, 10-, or 15-query samples which adhere to standard practices in FSL benchmarks [39]. To accommodate our dataset size and computational limitations, we train each model for 100 tasks per epoch and validate on 20 tasks. During testing, we assess performance using 200 tasks with the model state of the best validation result. We also use the same random seeds to train both models by controlling for random initialization to ensure consistency and reliability of results across experiments.

4) Optimization: During the training of our models, we utilized the SGD optimizer with Nesterov momentum and a weight decay of 0.0005. Initially, the learning rate was set to 0.1, and we employed the MultiStepLR scheduling technique to adjust the learning rate during training. Specifically, we trained the models for 50 epochs and applied a decrease factor of 10 to the learning rate after reaching 60% and 80% of the total training progress. However, to improve the training process further, we extended the training duration from 50 to 100 epochs while maintaining the same decrease factor for the learning rate. This extension allowed for more training iterations, potentially enhancing the model's performance. The selection of these parameter values was influenced by the research conducted by Laenen and Bertinetto [83], which provided valuable insights into effective training strategies.

III. RESULTS AND DISCUSSION

In this section, we compare the findings of other researchers who utilized different benchmark datasets, enabling a broader assessment of our approach's effectiveness and generalizability. Additionally, we present the results of our experiments, comparing the performance of PNs and MNs across different training settings, including varying ways, shots, and query samples.

A. Comparative Evaluation of FSL Using Novel Dataset

We conducted a comparative evaluation to assess the effectiveness of our proposed method for skin type classification, an image classification task. Leveraging the PNs and matching networks MNs methods, originally evaluated on the Omniglot and miniImageNet datasets, we evaluated their performance on our novel skin type dataset [7], [8].

Our proposed method achieved competitive performance on the skin type dataset compared to the benchmark datasets for both PNs and MNs. Despite the differences in content and characteristics, our method demonstrated promising accuracy, showcasing its robustness and effectiveness in skin type classification.

These results in Table 2 validate the generalizability of PNs and MNs methods, originally designed for image classification tasks on Omniglot and *mini*ImageNet, to the skin type classification task. The high accuracy achieved on the MFSTD dataset highlights the adaptability of our approach for diverse image classification tasks, including skin type classification.

In summary, our comparative evaluation confirms the robustness and effectiveness of our proposed method for skin type classification. By leveraging PNs' and MNs' methods and comparing the results with benchmark datasets, we establish the generalizability and performance of our approach. Table 2 provides a concise overview of the classification accuracies, reinforcing the efficacy of our method for image classification tasks, specifically in the context of skin type classification.

TABLE II
COMPARATIVE RESULTS OF OUR SKIN TYPE CLASSIFICATION USING THE MFSTD DATASET ALONGSIDE THE RESULTS REPORTED ON THE OMNIGLOT AND MINIIMAGENET DATASETS FOR PNS AND MNS.

Model	Omniglot 5-way Accuracy		*mini*ImageNet 5-way Accuracy		MFSTD (Ours) 4-way Accuracy	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PNs	98.80%	99.70%	41.20%	56.20%	89.66%	91.52%
MNs	98.10%	98.90%	46.60%	60.00%	58.25%	31.67%

B. Comparison of PNs and MNs

In this section, we present a detailed comparison between PNs and MNs regarding their performance and characteristics. The objective is to highlight the strengths and weaknesses of each model and provide insights into their suitability for our specific use case.

1) Performance on Few-Shot Classification: We evaluate the performance of PNs and MNs on few-shot classification tasks across various shot settings and report the corresponding classification accuracies. Based on Table 3, we can draw several key findings from the results obtained after 50 and 100 training epochs. PNs consistently outperform MNs in terms of accuracy across different shot settings. PNs demonstrate increasing accuracy as the number of shots increases, indicating their ability to leverage additional training examples per class. In contrast, MNs exhibit varying accuracy, achieving high accuracy in low-shot scenarios but struggling to maintain comparable performance in the 10-shot scenario.

After 50 training epochs, PNs achieve impressive accuracies ranging from 77.48% to 94.93% across different shot settings. This highlights the effectiveness of PNs in FSL tasks, showcasing their ability to generalize and make accurate predictions even with limited training examples.

MNs, on the other hand, show accuracy ranging from 45.02% to 90.27% after 50 epochs of training. While MNs perform well in low-shot scenarios, their accuracy drops significantly in the 10-shot scenario, suggesting challenges in generalization when faced with a more significant number of training examples per class.

Continuing the training for 100 epochs further improves the performance of both PNs and MNs. PNs consistently exhibit higher accuracies compared to MNs across different shot settings. PNs achieve accuracies ranging from 82.68% to 95.78% after 100 epochs, demonstrating their robustness and capability to handle increased shot settings.

MNs, on the other hand, still struggle to maintain comparable accuracy in the 10-shot scenario even after 100 epochs of training. Their accuracy ranges from 31.67% to 90.13%, indicating limitations in effectively generalizing with more training examples per class.

The result shows the significance of PNs as a powerful approach for FSL. The consistent superiority of PNs over MNs, particularly in scenarios with more training examples, demonstrates their potential for accurate classification in various real-world applications. Furthermore, the observed improvements with longer training durations emphasize the importance of model optimization and suggest the potential for further enhancements through extended training.

2) Sensitivity to Distance Metrics: We also investigate the impact of different distance metrics on the performance of PNs and MNs. MNs commonly employ cosine distance, which is bounded between -1 and 1. This bounded nature limits the attention function's ability to strongly emphasize a specific sample in the support set, potentially leading to slower convergence [84]. In contrast, PNs utilize unbounded Euclidean distance, which provides a broader range of values. Euclidean distance allows for faster convergence and better differentiation between samples, enabling the model to distinguish between fine-grained differences.

TABLE III

FEW-SHOT CLASSIFICATION ACCURACIES OF PNS AND MNS FOR COMPARISON. THE ACCURACIES ARE COMPUTED OVER 200 TEST EPISODES, WITH A FIXED 15 QUERIES, AND EVALUATED AT BOTH 50 EPOCHS AND 100 EPOCHS. THE BEST RESULT FOR EACH CATEGORY IS HIGHLIGHTED IN BOLD.

Model	2-way Accuracy 1-shot	2-way Accuracy 5-shot	2-way Accuracy 10-shot	3-way Accuracy 1-shot	3-way Accuracy 5-shot	3-way Accuracy 10-shot	4-way Accuracy 1-shot	4-way Accuracy 5-shot	4-way Accuracy 10-shot
PNs (50 epochs)	77.48%	**93.70%**	**94.93%**	85.49%	**93.38%**	**93.52%**	81.33%	**89.01%**	**90.55%**
MNs (50 epochs)	**82.32%**	90.27%	67.97%	**86.88%**	79.76%	52.76%	**87.08%**	45.02%	52.98%
PNs (100 epochs)	88.53%	**94.18%**	**95.78%**	87.48%	**93.54%**	**93.93%**	82.68%	**89.66%**	**91.52%**
MNs (100 epochs)	**90.13%**	63.95%	72.68%	**89.24%**	90.11%	62.89%	**87.61%**	58.25%	31.67%

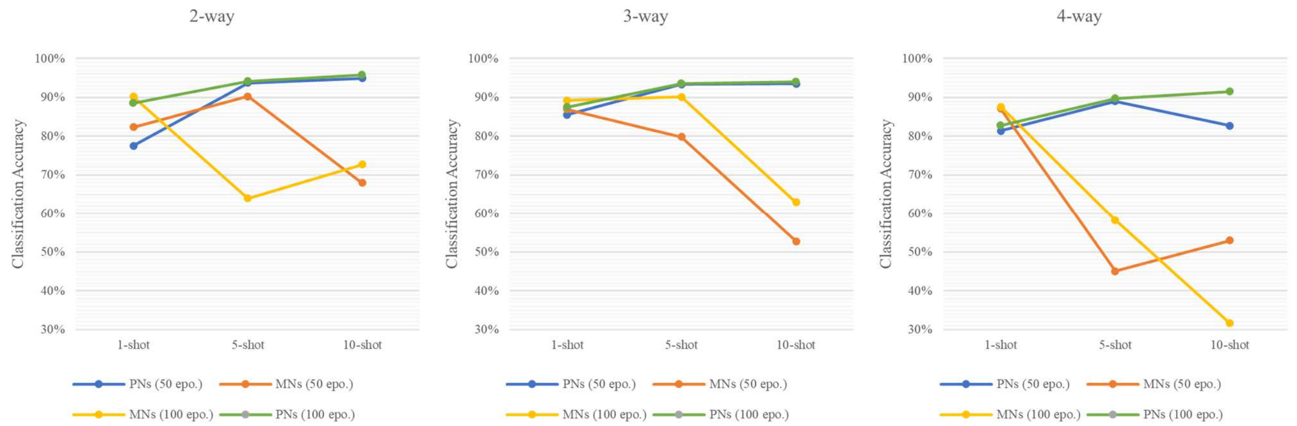


Fig. 9 Comparison graph illustrates the effect of classes on 2-, 3-, and 4-way classification accuracy for PNs and MNs.

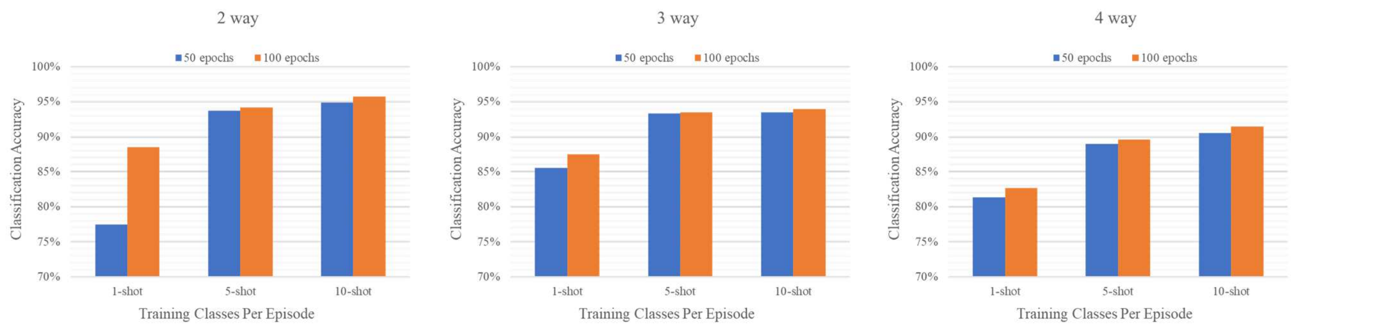


Fig. 10 Impact of training epochs on the classification accuracy of PNs in different few-shot classification scenarios. The results include 2-way, 3-way, and 4- way classification tasks with different shots per class while keeping the number of query samples constant at 15.

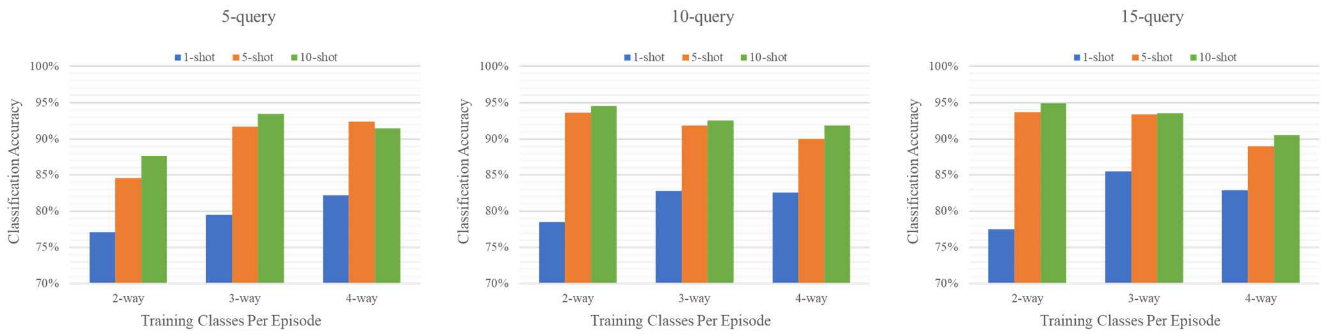


Fig. 11 Impact of the number of ways, shots, and queries on the accuracy of FSL models. The accuracy is evaluated with a fixed number of 50 training epochs.

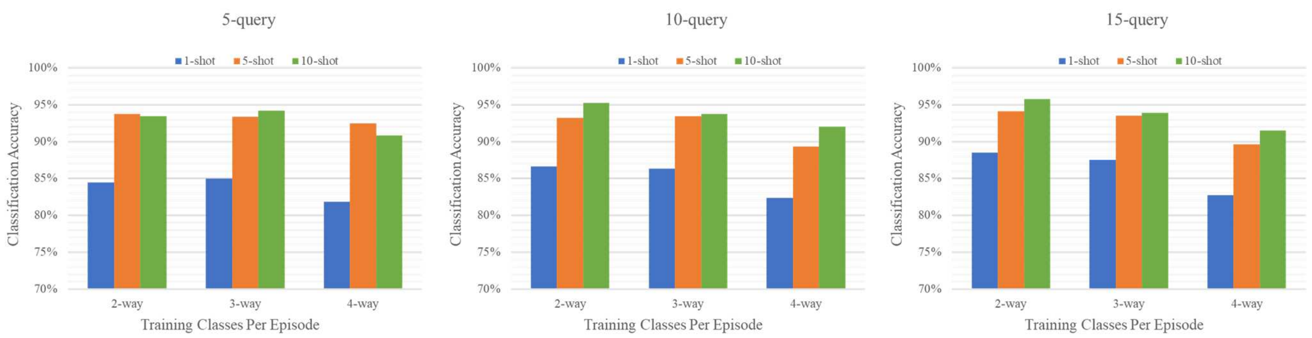


Fig. 12 Impact of the number of ways, shots, and queries on the accuracy of FSL models. The accuracy is evaluated with a fixed number of 100 training epochs.

These findings highlight the sensitivity of the models to the choice of a distance metric, suggesting that Euclidean distance used by PNs can be advantageous for faster convergence and improved discrimination between samples compared to cosine distance commonly employed by MNs.

3) Generalization and robustness: PNs demonstrate strong generalization abilities, performing well even with limited training examples per class. On the other hand, MNs may face challenges in effectively generalizing when confronted with a more significant number of training examples per class, leading to a decrease in accuracy. Hence, the choice between PNs and MNs should consider the specific requirements of the task, particularly the availability of training data.

4) Model complexity and training efficiency: PNs have a relatively more straightforward architecture involving calculating prototype vectors for each class. On the other hand, MNs require the generation of attention weights for each sample in the support set, which can be computationally intensive. PNs also exhibit faster convergence compared to MNs, as discussed earlier.

These results provide insights into the performance, sensitivity to distance metrics, generalization abilities, and model complexity of PNs and MNs. Researchers and practitioners can utilize this information to make informed decisions when selecting the appropriate model for their specific few-shot classification tasks, considering factors such as available training data and desired trade-offs between accuracy and computational efficiency.

C. Performance of PNs on Our Dataset

In this section, we conduct an extensive performance analysis of PNs on our dataset. We emphasize PNs as they are more suitable than MNs in our research context. Additionally, we will discuss the impact of training epochs, the combined influence of way, shot, and query on performance, and the entire test classification accuracy comparison between PNs and MNs.

1) Impact of training epochs: We first investigate the impact of the number of training epochs on the performance of PNs. Fig. 10 shows the classification accuracy results for 50 and 100 epochs. It can be observed that increasing the number of epochs generally leads to improved classification accuracy.

The accuracy of the 2-way classification scenario significantly improves with more training epochs. For 1-shot 15-query tasks, the accuracy increases from 77.48% after 50 epochs to 88.53% after 100 epochs. This improvement highlights the model's ability to learn and generalize patterns with additional training iterations. However, for 5-shot and 10-shot tasks, accuracy remains consistently high throughout the 50 and 100 epochs, with scores ranging from 93.70% to 95.78%. These results suggest that even a limited number of training epochs is sufficient for the model to perform strongly in distinguishing between two classes.

Moving on to the 3-way classification, we find that the impact of training epochs is less pronounced. The accuracy for 1-shot 15-query tasks shows a modest improvement from 85.49% to 87.48% as the number of epochs increases from 50 to 100. The accuracy remains relatively stable for 5- and 10-shot tasks, ranging from 93.38% to 93.93%. These findings indicate that additional training iterations have a limited impact on the model's performance in discriminating among three classes.

The accuracy in the 4-way classification scenario exhibits minor improvements with more training epochs. For 1-shot 15-query tasks, the accuracy ranges from 81.33% to 82.68% across 50 and 100 epochs. The model's performance remains relatively stable for 5-shot tasks, with accuracy scores ranging from 89.01% to 89.66%. However, for 10-shot tasks, accuracy shows a moderate improvement from 90.55% to 91.52% with additional training epochs.

Overall, the analysis reveals that the impact of training epochs on classification accuracy varies depending on the number of classes in the few-shot learning scenario. While increasing the number of epochs generally improves accuracy, the magnitude of improvement differs across different classification scenarios. Notably, 2- and 3-way classifications tend to achieve higher accuracy than 4-way classifications, indicating the increasing difficulty of discriminating between larger classes.

D. The Combined Impact of Way, Shot, Query on Performance

The findings from our analysis reveal exciting patterns regarding the impact of way, as well as the number of shots and query images. Analyzing the results, we observed specific patterns and trends in the accuracy based on these factors. We noticed a lower initial accuracy when examining scenarios of 50 epochs with 5 query images. This can be attributed to the limited number of instances for the model to make predictions. A smaller data pool hinders the model's ability to accurately discern the correct class labels. However, as we increased the number of classes from 2- to 3-way, we observed a subsequent increase in accuracy. This suggests that the model benefits from the additional discrimination patterns introduced by the extra class. However, the trend reversed when moving to 4-way, with a drop in accuracy. This could be due to the increased complexity of the task, exceeding the model's capacity to differentiate between a more significant number of classes accurately.

On the other hand, scenarios with 10 and 15 query images exhibited a higher initial accuracy. This can be attributed to the more significant number of instances the model can learn from. With more diverse sample data, the model can capture more representative patterns and improve its initial performance. However, as we increased the number of classes, we consistently observed a drop in accuracy. This suggests that as the task becomes more complex, with more classes to discriminate between, the model faces challenges in accurately distinguishing between them.

The trends observed in the accuracy across different shot configurations are also noteworthy. Generally, an increase in the number of shots led to improved accuracy, as shown in Fig. 11 and Fig. 12. This can be attributed to the model having more examples to learn from during training. The additional shots provide more information and help the model better understand the distinguishing features of each class.

However, it is essential to note that the impact of shots is intertwined with other factors, such as the number of query images and classes. Therefore, finding the optimal balance between these variables is crucial to achieving the highest accuracy in few-shot learning. These findings highlight the complex dynamics between the number of classes, shots, and query images in FSL scenarios. The relationships among these factors are interconnected, and their influence on accuracy can vary based on the specific configuration.

E. Complete test classification accuracy of PNs and MNs

The findings from our analysis reveal exciting patterns regarding the impact of way, as well as the number of shots and query images. Analyzing the results, we observed specific patterns and comprehensively compared our dataset's complete test classification accuracy between PNs and MNs. By evaluating the performance of both models across various classification tasks, we determine the superior model in terms of accuracy and overall performance. This analysis validates our earlier assertion that PNs are more suitable than MNs for our specific research domain. The detailed results of PNs and MNs can be accessed in Appendix B and C, respectively. Based on the results obtained from our experiments, the best classification performance on the dataset was observed in the 2-way classification scenario with 10 query images of PNs. In this configuration, the model achieved an accuracy of 94.93% after 50 epochs and improved to 95.78% after 100 epochs of training.

This result suggests that combining a smaller number of classes (2-way) and a relatively more significant number of query images (10) yielded the highest accuracy in the few-shot classification task. The model's ability to effectively discriminate between a limited number of classes and the availability of diverse instances for learning and inference likely contributed to this improved performance. It is important to note that these results are specific to our study's dataset and experimental setup. Other factors, such as the dataset's nature, the classification task's complexity, and the choice of model architecture, may influence the optimal configuration for achieving the best performance.

IV. CONCLUSION

While there have been great advancements in the fields of face recognition [85]–[91] and facial expressions [92]–[98], not much attention has been given to using machine learning techniques for analyzing different skin types. This paper introduced a method called PNs for skin type classification using few-shot learning. PNs achieved state-of- the-art results of 95.78% accuracy in this task. Our approach focuses on the fundamental aspects of architecture and employs episodic training, resulting in a simple yet efficient solution. We also explored the potential of PNs in handling skin typing systems with a more significant number of classes, going beyond the typical few-shot setting.

Looking ahead, there are several promising research directions. We aim to investigate the feasibility of using PNs for skin type classification with more classes and alternative skin typing systems. Additionally, comparing PNs with other few-shot learning models like Relation Networks or SimpleShot would provide valuable insights into their strengths and weaknesses [99], [100], [101]. It would also be beneficial to explore the compatibility of PNs with different ResNet backbone architectures to enhance their versatility.

To validate the practicality of PNs, we plan to collaborate with dermatologists and conduct validation studies in clinical settings. This collaboration will help us understand the real-world applicability of PNs in skin type classification and improve their accuracy and reliability.

In conclusion, our study demonstrates the promise of PNs for FSL in skin type classification. With their simplicity, efficiency, and competitive performance, PNs offer a compelling approach. By further exploring different variations, testing on alternative skin typing systems, comparing with other models, and collaborating with domain experts, we can advance the field of skin type classification and develop more accurate and robust solutions.

REFERENCES

[1] R. Oliveira, J. Ferreira, L. F. Azevedo, and I. F. Almeida, "An Overview of Methods to Characterize Skin Type: Focus on Visual Rating Scales and Self-Report Instruments," *Cosmetics 2023, Vol. 10, Page 14*, vol. 10, no. 1, p. 14, Jan. 2023, doi:10.3390/COSMETICS10010014.

[2] A. V. Rawlings, "Ethnic skin types: Are there differences in skin structure and function?," *Int J Cosmet Sci*, vol. 28, no. 2, pp. 79–93, 2006, doi:10.1111/j.1467-2494.2006.00302.x.

[3] I. L. Chan, S. Cohen, M. G. da Cunha, and L. C. Maluf, "Characteristics and management of Asian skin," *Int J Dermatol*, vol. 58, no. 2, pp. 131–143, 2019, doi:10.1111/ijd.14153.

[4] B. Jothishankar and S. L. Stein, "Impact of skin color and ethnicity," *Clin Dermatol*, vol. 37, no. 5, pp. 418–429, 2019, doi:10.1016/j.clindermatol.2019.07.009.

[5] C. F. Goh, N. Mohamed Faisal, and F. N. Ismail, "Facial Skin Biophysical Profile of Women in Malaysia: Significance of Facial Skincare Product Use," *Skin Pharmacol Physiol*, vol. 34, no. 6, pp. 351–362, 2021, doi:10.1159/000514995.

[6] C. F. Goh, "Diversity of Asian skin: A review on skin biophysical properties," Exp Dermatol, 2023, doi:10.1111/exd.14959.

[7] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in Advances in Neural Information Processing Systems, 2017, pp. 4078–4088.

[8] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in Advances in Neural Information Processing Systems, 2016, pp. 3637–3645.

[9] P. Tschandl et al., "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," Lancet Oncol, vol. 20, no. 7, pp. 938–947, 2019, doi:10.1016/S1470-2045(19)30333-X.

[10] H. Ibrahim, X. Liu, N. Zariffa, A. D. Morris, and A. K. Denniston, "Health data poverty: an assailable barrier to equitable digital health care," Lancet Digit Health, vol. 3, no. 4, pp. e260–e265, 2021, doi:10.1016/S2589-7500(20)30317-4.

[11] S. Burgin, N. C. Dlova, and L. A. Goldsmith, "Dermatological education for the 21st century: prioritizing diversity," British Journal of Dermatology, vol. 184, no. 4, pp. 740–741, 2021, doi:10.1111/bjd.19663.

[12] A. Bouslimani et al., "The impact of skin care products on skin chemistry and microbiome dynamics," BMC Biol, vol. 17, no. 1, 2019, doi:10.1186/s12915-019-0660-6.

[13] L. Tizek, M. C. Schielein, F. Seifert, T. Biedermann, A. Böhner, and A. Zink, "Skin diseases are more common than we think: screening results of an unreferred population at the Munich Oktoberfest," Journal of the European Academy of Dermatology and Venereology, vol. 33, no. 7, pp. 1421–1428, 2019, doi:10.1111/jdv.15494.

[14] L. Zhang et al., "The Impact of Routine Skin Care on the Quality of Life," Cosmetics, vol. 7, no. 3, 2020, doi:10.3390/cosmetics7030059.

[15] M. Nagae, T. Mitsutake, and M. Sakamoto, "Impact of skin care on body image of aging people: A quasi-randomized pilot trial," Heliyon, vol. 9, no. 2, 2023, doi:10.1016/j.heliyon.2023.e13230.

[16] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115– 118, Feb. 2017, doi:10.1038/nature21056.

[17] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin Lesion Classification Using Hybrid Deep Neural Networks," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, May 2019, pp. 1229–1233. doi:10.1109/ICASSP.2019.8683352.

[18] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention Residual Learning for Skin Lesion Classification," IEEE Trans Med Imaging, vol. 38, no. 9, pp. 2092–2103, Sep. 2019, doi: 10.1109/TMI.2019.2893944.

[19] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," Artif Intell Med, vol. 102, 2020, doi:10.1016/j.artmed.2019.101756.

[20] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," Sensors, vol. 21, no. 8, p. 2852, Apr. 2021, doi:10.3390/s21082852.

[21] A. Esteva et al., "Deep learning-enabled medical computer vision," NPJ Digit Med, vol. 4, no. 1, p. 5, Jan. 2021, doi:10.1038/s41746- 020-00376-2.

[22] D. Sonntag, F. Nunnari, and H.-J. Profitlich, "The Skincare project, an interactive deep learning system for differential diagnosis of malignant skin lesions. Technical Report," May 2020, [Online]. Available: http://arxiv.org/abs/2005.09448

[23] W. Zhu, H. Liao, W. Li, W. Li, and J. Luo, Alleviating the Incompatibility Between Cross Entropy Loss and Episode Training for Few-Shot Skin Disease Classification, vol. 12266 LNCS. 2020. doi: 10.1007/978-3-030-59725-2_32.

[24] Y. Wang et al., "Cross-Domain Few-Shot Learning for Rare-Disease Skin Lesion Segmentation," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2022, pp. 1086–1090. doi:10.1109/ICASSP43922.2022.9746791.

[25] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," BMC Med, vol. 17, no. 1, 2019, doi:10.1186/s12916-019-1426-2.

[26] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J Big Data, vol. 8, no. 1, p. 53, Mar. 2021, doi:10.1186/s40537-021- 00444-8.

[27] L. Baumann, "Understanding and Treating Various Skin Types: The Baumann Skin Type Indicator," Dermatol Clin, vol. 26, no. 3, pp. 359–373, Jul. 2008, doi: 10.1016/J.DET.2008.03.007.
[28] L. S. Baumann et al., "A Validated Questionnaire for Quantifying Skin Oiliness," Journal of Cosmetics, Dermatological Sciences and Applications, vol. 4, no. 2, pp. 78–84, Mar. 2014, doi:10.4236/JCDSA.2014.42012.
[29] S. I. Cho, D. Kim, H. Lee, T. T. Um, and H. Kim, "Explore highly relevant questions in the Baumann skin type questionnaire through the digital skin analyzer: A retrospective single-center study in South Korea," J Cosmet Dermatol, vol. 22, no. 11, pp. 3159–3167, 2023, doi:10.1111/jocd.15820.
[30] R. Efata, W. I. Loka, N. Wijaya, and D. Suhartono, "Facial Skin Type Prediction Based on Baumann Skin Type Solutions Theory Using Machine Learning," TEM Journal, vol. 12, no. 1, pp. 96–103, 2023, doi:10.18421/TEM121-13.
[31] Y. B. Lee et al., "Baumann skin type in the Korean male population," Ann Dermatol, vol. 31, no. 6, pp. 621–630, 2019, doi:10.5021/ad.2019.31.6.621.
[32] S. Kanezawa, Y.-B. Zhu, and Q. Wang, "Correlation between Chinese Medicine Constitution and Skin Types: A Study on 187 Japanese Women," Chin J Integr Med, vol. 26, no. 3, pp. 174–179, 2020, doi:10.1007/s11655-019-2709-3.
[33] J. Y. Hong, S. J. Park, S. J. Seo, and K. Y. Park, "Oily sensitive skin: A review of management options," J Cosmet Dermatol, vol. 19, no. 5, pp. 1016–1020, 2020, doi:10.1111/jocd.13347.
[34] P. Magin, D. Pond, W. Smith, S. Goode, and N. Paterson, "Reliability of skin-type self-assessment: agreement of adolescents' repeated Fitzpatrick skin phototype classification ratings during a cohort study," J Eur Acad Dermatol Venereol, vol. 26, no. 11, pp. 1396–1399, Nov. 2012, doi:10.1111/J.1468-3083.2011.04298.X.
[35] V. Gupta and V. K. Sharma, "Skin typing: Fitzpatrick grading and others," Clin Dermatol, vol. 37, no. 5, pp. 430–436, 2019, doi:10.1016/j.clindermatol.2019.07.010.
[36] O. R. Ware, J. E. Dawson, M. M. Shinohara, and S. C. Taylor, "Racial limitations of fitzpatrick skin type," Cutis, vol. 105, no. 2, pp. 77–80, 2020.
[37] M. Fors, P. González, C. Viada, K. Falcon, and S. Palacios, "Validity of the Fitzpatrick Skin Phototype Classification in Ecuador," Adv Skin Wound Care, vol. 33, no. 12, pp. 1–5, Dec. 2020, doi:10.1097/01.ASW.0000721168.40561.a3.
[38] W. E. Robert, "The Roberts Skin Type Classification System," Journal of Drugs in Dermatology, vol. 7, no. 5, pp. 452–456, May 2008, Accessed: Nov. 02, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18505137/
[39] A. Parnami and M. Lee, "Learning from Few Examples: A Summary of Approaches to Few-Shot Learning," Mar. 2022, Accessed: Nov. 02, 2023. [Online]. Available: http://arxiv.org/abs/2203.04291
[40] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-shot Learning," ACM Comput Surv, vol. 53, no. 3, 2020, doi:10.1145/3386252.
[41] X. Zhang, C. Wang, Y. Tang, Z. Zhou, and X. Lu, A Survey of Few-Shot Learning and Its Application in Industrial Object Detection Tasks, vol. 880 LNEE. 2022. doi:10.1007/978-981-19-0572-8_81.
[42] K.-L. Zhao, X.-L. Jin, and Y.-Z. Wang, "Survey on Few-shot Learning" Ruan Jian Xue Bao/Journal of Software, vol. 32, no. 2, pp. 349–369, 2021, doi: 10.13328/j.cnki.jos.006138.
[43] E. Bennequin, "Few-Shot Image Classification with Meta- Learning." Accessed: Nov. 02, 2023. [Online]. Available: https://www.sicara.fr/blog-technique/few-shot-image-classification-meta-learning
[44] J. Vanschoren, "Meta-Learning," in Automated Machine Learning: Methods, Systems, Challenges, L. and V. J. Hutter Frank and Kotthoff, Ed., Cham: Springer International Publishing, 2019, pp. 35–61. doi:10.1007/978-3-030-05318-5_2.
[45] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-Learning in Neural Networks: A Survey," IEEE Trans Pattern Anal Mach Intell, vol. 44, no. 9, pp. 5149–5169, 2022, doi:10.1109/TPAMI.2021.3079209.
[46] A. Rivolli, L. P. F. Garcia, C. Soares, J. Vanschoren, and A. C. P. L. F. de Carvalho, "Meta-features for meta-learning," Knowl Based Syst, vol. 240, 2022, doi:10.1016/j.knosys.2021.108101.
[47] K. Lee, S. Maji, A. Ravichandran, and S. Soatto, "Meta-learning with differentiable convex optimization," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, pp. 10649–10657. doi:9910.1109/CVPR.2019.01091.
[48] A. Antoniou, A. Storkey, and H. Edwards, "How to train your MAML," in 7th International Conference on Learning Representations, ICLR 2019, 2019.
[49] K. He, N. Pu, M. Lao, and M. S. Lew, "Few-shot and meta-learning methods for image understanding: a survey," Int J Multimed Inf Retr, vol. 12, no. 2, 2023, doi:10.1007/s13735-023-00279-4.
[50] Y. Tian, X. Zhao, W. Huang, "Meta-learning approaches for learning-to-learn in deep learning: A survey," Neurocomputing, vol.494, pp. 203–223, 2022, doi: 10.1016/j.neucom.2022.04.078.
[51] M. Huisman, J. N. van Rijn, and A. Plaat, "A survey of deep meta-learning," Artif Intell Rev, vol. 54, no. 6, pp. 4483–4541, 2021, doi: 10.1007/s10462-021-10004-4.
[52] D. Wang, Y. Cheng, M. Yu, X. Guo, and T. Zhang, "A hybrid approach with optimization-based and metric-based meta-learner for few-shot learning," Neurocomputing, vol. 349, pp. 202–211, 2019, doi: 10.1016/j.neucom.2019.03.085.
[53] D. Gupta and K. K. Shukla, A Study on Metric-Based and Initialization-Based Methods for Few-Shot Image Classification, vol. 547. 2023. doi:10.1007/978-981-19-6525-8_4.
[54] P. H. Sulibhavi et al., "A Study On Meta Learning Optimization Techniques," in 2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022, 2022. doi: 10.1109/I2CT54291.2022.9824942.
[55] X. Li, X. Yang, Z. Ma, and J.-H. Xue, "Deep metric learning for few-shot image classification: A Review of recent developments," Pattern Recognit, vol. 138, 2023, doi: 10.1016/j.patcog.2023.109381.
[56] D. Jung, D. Kang, S. Kwak, and M. Cho, Few-shot Metric Learning: Online Adaptation of Embedding for Retrieval, vol. 13845 LNCS. 2023. doi: 10.1007/978-3-031-26348-4_4.
[57] S. Kim, M. Seo, I. Laptev, M. Cho, and S. Kwak, "Deep metric learning beyond binary supervision," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, pp. 2283–2292. doi: 10.1109/CVPR.2019.00239.
[58] J. Tang, D. Li, and Y. Tian, "Image classification with multi-view multi-instance metric learning," Expert Syst Appl, vol. 189, 2022, doi: 10.1016/j.eswa.2021.116117.
[59] Y. Li, "Supervised Few-Shot Image Segmentation with Deep Metric Learning," in Proceedings - 2021 International Conference on Electronic Information Technology and Smart Agriculture, ICEITSA 2021, 2021, pp. 431–434. doi: 10.1109/ICEITSA54226.2021.00088.
[60] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D, vol. 404, 2020, doi: 10.1016/j.physd.2019.132306.
[61] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-Learning with Memory-Augmented Neural Networks," in 33rd International Conference on Machine Learning, ICML 2016, 2016, pp. 2740–2751.
[62] S. Deng, N. Zhang, J. Kang, Y. Zhang, W. Zhang, and H. Chen, "Meta-learning with dynamic-memory-based prototypical network for few-shot event detection," in WSDM 2020 - Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 151–159. doi: 10.1145/3336191.3371796.
[63] Y. Pan, T. Yao, Y. Li, Y. Wang, C.-W. Ngo, and T. Mei, "Transferrable prototypical networks for unsupervised domain adaptation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, pp. 2234–2242. doi: 10.1109/CVPR.2019.00234.
[64] T. Gao, X. Han, Z. Liu, and M. Sun, "Hybrid attention-based prototypical networks for noisy few-shot relation classification," in 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, 2019, pp. 6407–6414.
[65] T. Ko, Y. Chen, and Q. Li, "Prototypical Networks for Small Footprint Text-Independent Speaker Verification," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, doi:10.1109/icassp40776.2020.9054471.
[66] M. Ren et al., "Meta-learning for semi-supervised few-shot classification," in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
[67] G. Song, Z. Tao, X. Huang, G. Cao, W. Liu, and L. Yang, "Hybrid Attention-Based Prototypical Network for Unfamiliar Restaurant Food Image Few-Shot Recognition," IEEE Access, vol. 8, pp. 14893–14900, 2020, doi: 10.1109/access.2020.2964836.
[68] X. Li, Z. Cao, L. Zhao, and J. Jiang, "ALPN: Active-Learning- Based Prototypical Network for Few-Shot Hyperspectral Imagery

Classification," IEEE Geoscience and Remote Sensing Letters, vol. 19, 2022, doi: 10.1109/LGRS.2021.3101495.
[69] S. Huang, X. Zeng, S. Wu, Z. Yu, M. Azzam, and H.-S. Wong, "Behavior regularized prototypical networks for semi-supervised few-shot image classification," Pattern Recognit, vol. 112, 2021, doi:10.1016/j.patcog.2020.107765.
[70] B. Xi et al., "Deep Prototypical Networks with Hybrid Residual Attention for Hyperspectral Image Classification," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 13, pp. 3683–3700, 2020, doi: 10.1109/JSTARS.2020.3004973.
[71] H. Tang, Y. Li, X. Han, Q. Huang, and W. Xie, "A Spatial-Spectral Prototypical Network for Hyperspectral Remote Sensing Image," IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 1, pp. 167–171, 2020, doi: 10.1109/LGRS.2019.2916083.
[72] Z. Ji, X. Chai, Y. Yu, Y. Pang, and Z. Zhang, "Improved prototypical networks for few-Shot learning," Pattern Recognit Lett, vol. 140, pp. 81–87, 2020, doi: 10.1016/j.patrec.2020.07.015.
[73] C. Zhang, J. Yue, and Q. Qin, "Global prototypical network for few-shot hyperspectral image classification," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 13, pp. 4748–4759, 2020, doi:10.1109/JSTARS.2020.3017544.
[74] V. Prabhu, A. Kannan, M. Ravuri, M. Chablani, D. Sontag, and X. Amatriain, "Few-shot learning for dermatological disease diagnosis," in Meta Learning With Medical Imaging and Health Informatics Applications, H. Van Nguyen, R. Summers, and R. Chellappa, Eds., in The MICCAI Society book Series. , Elsevier, 2023, pp. 235–252. doi: 10.1016/B978-0-32-399851-2.00022-3.
[75] K. Mahajan, M. Sharma, and L. Vig, "Meta-DermDiagnosis: Few-Shot Skin Disease Identification using Meta-Learning," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Jun. 2020, pp. 3142– 3151. doi: 10.1109/CVPRW50498.2020.00373.
[76] X.-J. Liu, K. Li, H. Luan, W. Wang, and Z. Chen, "Few-shot learning for skin lesion image classification," Multimed Tools Appl, vol. 81, no. 4, pp. 4979–4990, Feb. 2022, doi: 10.1007/s11042-021- 11472-0.
[77] Q. Lu, "Asian Cosmetics in Global Market: A Comparative Study of Internationalization of Japanese, Korean, and Chinese Companies," Harvard Extension School, 2020. Accessed: Nov. 02, 2023. [Online]. Available: https://dash.harvard.edu/handle/1/37365641
[78] L. Baumann, "Baumann Skin Type 10: ORNT | Best Skin Type – Skin Type Solutions." Accessed: Nov. 02, 2023. [Online]. Available: https://skintypesolutions.com/blogs/skincare/baumann- skin-type-10-ornt?_pos=1&_sid=33db792ba&_ss=r
[79] L. Baumann, "The baumann skin-type indicator: A novel approach to understanding skin type," in Handbook of Cosmetic Science and Technology, Third Edition, CRC Press, 2009, pp. 29–40.
[80] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
[81] S. Ravi and H. Larochelle, "Optimization as a Model for Few-Shot Learning," in International Conference on Learning Representations, 2017. [Online]. Available: https://openreview.net/forum?id=rJY0-Kcll.
[82] G. Ghiasi, T.-Y. Lin, and Q. V Le, "DropBlock: A regularization method for convolutional networks," in Advances in Neural Information Processing Systems, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/7edcfb2d 8f6a659ef4cd1e6c9b6d7079-Paper.pdf
[83] S. Laenen and L. Bertinetto, "On Episodes, Prototypical Networks, and Few-Shot Learning," in Advances in Neural Information Processing Systems, 2021, pp. 24581–24592.
[84] O. Knagg, "Advances in few-shot learning: reproducing results in PyTorch | by Oscar Knagg | Towards Data Science." Accessed: Nov. 02, 2023. [Online]. Available: https://towardsdatascience.com/advances-in-few-shot-learning-reproducing-results-in-pytorch-aba70dee541d
[85] L. Li, X. Mu, S. Li, and H. Peng, "A Review of Face Recognition Technology," IEEE Access, vol. 8, pp. 139110–139120, 2020, doi:10.1109/access.2020.3011028.
[86] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, pp. 4685–4694. doi: 10.1109/CVPR.2019.00482.
[87] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," IEEE Trans Pattern Anal Mach Intell, vol. 41, no. 1, pp. 121–135, 2019, doi:10.1109/TPAMI.2017.2781233.
[88] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439–2450, 2019, doi: 10.1109/TIP.2018.2886767.
[89] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," Measurement (Lond), vol. 167, 2021, doi:10.1016/j.measurement.2020.108288.
[90] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131–148, 2020, doi:10.1016/j.inffus.2020.06.014.
[91] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," Artif Intell Rev, vol. 52, no. 2, pp. 927–948, 2019, doi:10.1007/s10462-018-9650-2.
[92] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," Journal of Informatics and Web Engineering, vol. 1, no. 2, pp. 45–54, Sep. 2022, doi:10.33093/jiwe.2022.1.2.4.
[93] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements," Psychological Science in the Public Interest, vol. 20, no. 1, pp. 1–68, 2019, doi:10.1177/1529100619832930.
[94] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Trans Affect Comput, vol. 10, no. 1, pp. 18–31, 2019, doi:10.1109/TAFFC.2017.2740923.
[95] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition," IEEE Transactions on Image Processing, vol. 29, pp. 4057–4069, 2020, doi:10.1109/TIP.2019.2956143.
[96] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 356–370, 2019, doi: 10.1109/TIP.2018.2868382.
[97] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Trans Affect Comput, vol. 13, no. 3, pp. 1195–1215, 2022, doi: 10.1109/TAFFC.2020.2981446.
[98] Y. Wu and Q. Ji, "Facial Landmark Detection: A Literature Survey," Int J Comput Vis, vol. 127, no. 2, pp. 115–142, 2019, doi: 10.1007/s11263-018-1097-z.
[99] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to Compare: Relation Network for Few-Shot Learning," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208. doi: 10.1109/CVPR.2018.00131.
[100] Y. Wang, W.-L. Chao, K. Q. Weinberger, and L. van der Maaten, "SimpleShot: Revisiting Nearest-Neighbor Classification for Few-Shot Learning," Nov. 2019, Accessed: Nov. 04, 2023. [Online]. Available: https://arxiv.org/abs/1911.04623v2
[101] K. Gao, B. Liu, X. Yu, J. Qin, P. Zhang, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," Remote Sens (Basel), vol. 12, no. 6, 2020, doi: 10.3390/rs12060923.

APPENDIX A

TABLE IV
MODIFIED BSTI QUESTIONNAIRE: SKIN CONDITION ASSESSMENT

Question
1. In photos, your skin tends to appear: a) Matte b) Mostly matte c) Sometimes, shiny d) I shine like a diamond e) I have never noticed anything
2. When you wake up in the morning, your skin feels: a) Tight or dry b) Comfortable c) Oily in T-zone d) Oily e) Not sure
3. After a shower, your face tends to feel: a) Tight or dry b) No particular sensation c) Slight sheen d) Oily e) I have never noticed anything
4. In the afternoon, my skin needs: a) Moisturizing all over b) A refreshing spritzes of facial spray c) Blotting or powdering on the forehead, nose, and chin d) Blotting or powder all over
5. How does it feel when you touch your skin? a) Rough and scaly b) Oily in places and dry in others c) Slick and greasy d) Irritated and angry
6. How would you describe your skin type? a) Dry b) Neutral/ Normal c) Combination d) Oily e) Not sure
7. I would describe the shine of my skin like this: a) Dull everywhere b) Shinny in my T-zone, but dull on my cheeks c) Bright like a diamond d) I get more stinging than shine
8. How does your skin feel after you wash your face? a) Itchy and dry b) Stripped of moisture c) Clean and great in my T-zone, but my cheeks are a little bit dried out d) Clean for now, but the oil is coming soon
9. Which most closely describes the look of your pores? a) Large and visible all over

b) Larger or medium and only visible in the T- zone
c) Medium-sized all over
d) Small and not easily noticed all over

10. When does your skin look red?
    a) More often than I care to admit
    b) Whenever and wherever I use new products
    c) Sometimes, but only around my cheeks
    d) Anytime I have blemishes
11. Pick the one that best describes your skin's relationship with pimples and blackheads.
    a) My blemishes are more likely to be broken capillaries or rashes
    b) They have their territory around my T-zone, and I have claimed the cheeks for my own
    c) I hate them, but they love me
    d) I would trade my skin flakiness and tightness for a few blackheads
12. Is your skin wrinkled?
    a) Yes
    b) No
13. If Yes, where do you have wrinkles? (You may choose more than one)
    a) Forehead
    b) Nose
    c) Cheeks
    d) Chin
14. What is your skin-pigmented problem? (You may choose more than one).
    a) Melasma
    b) Solar lentigos
    c) Ephelides
    d) Post-inflammatory hyperpigmentation
    e) None of the above
15. Do you require any skin-lightening products/ingredients?
    a) Yes
    b) No

APPENDIX B

TABLE V

SUMMARY OF THE CLASSIFICATION ACCURACY OF PNS ON THE TEST SET AFTER 50 AND 100 TRAINING EPOCHS. THE REPORTED ACCURACY VALUES ARE AVERAGED OVER 200 RANDOMLY GENERATED TEST EPISODES, WITH 95% CONFIDENCE INTERVALS SHOWN TO INDICATE THE RESULTS' CERTAINTY LEVEL. THE BEST RESULT FOR EACH CLASS IS HIGHLIGHTED IN BOLD.

	Train Episodes			**Classification Accuracy**	
Model	Way	Shot	Query	50 epochs	100 epochs
PNs	2	1	5	77.10 ± 5.82%	84.45 ± 5.02%
PNs	2	1	10	78.50 ± 5.69%	86.60 ± 4.72%
PNs	2	1	15	77.48 ± 5.79%	88.53 ± 4.42%
PNs	2	5	5	84.55 ± 5.01%	93.75 ± 3.35%
PNs	2	5	10	92.90 ± 3.56%	93.25 ± 3.48%
PNs	2	5	15	93.70 ± 3.37%	94.18 ± 3.24%
PNs	2	10	5	87.65 ± 4.56%	93.45 ± 3.43%
PNs	2	10	10	94.53 ± 3.15%	95.28 ± 2.94%
PNs	2	10	15	**94.93 ± 3.04%**	**95.78 ± 2.79%**
PNs	3	1	5	79.50 ± 5.60%	84.97 ± 4.95%
PNs	3	1	10	82.82 ± 5.23%	86.33 ± 4.76%
PNs	3	1	15	85.49 ± 4.88%	87.48 ± 4.59%
PNs	3	5	5	91.73 ± 3.82%	93.43 ± 3.43%
PNs	3	5	10	91.83 ± 3.80%	93.50 ± 3.42%
PNs	3	5	15	93.38 ± 3.45%	93.54 ± 3.41%
PNs	3	10	5	93.50 ± 3.42%	**94.23 ± 3.23%**
PNs	3	10	10	92.55 ± 3.64%	93.78 ± 3.35%
PNs	3	10	15	**93.52 ± 3.41%**	93.93 ± 3.31%
PNs	4	1	5	82.15 ± 5.31%	81.85 ± 5.34%
PNs	4	1	10	82.54 ± 5.26%	82.32 ± 5.29%
PNs	4	1	15	81.33 ± 5.40%	82.68 ± 5.24%
PNs	4	5	5	**92.40 ± 3.67%**	**92.47 ± 3.66%**
PNs	4	5	10	90.04 ± 4.15%	89.35 ± 4.28%
PNs	4	5	15	89.01 ± 4.33%	89.66 ± 4.22%
PNs	4	10	5	91.47 ± 3.87%	90.83 ± 4.00%
PNs	4	10	10	91.88 ± 3.79%	92.07 ± 3.74%
PNs	4	10	15	90.55 ± 4.05%	91.52 ± 3.86%

APPENDIX C

TABLE VI

SUMMARY OF THE CLASSIFICATION ACCURACY OF MNS ON THE TEST SET AFTER 50 AND 100 TRAINING EPOCHS. THE REPORTED ACCURACY VALUES ARE AVERAGED OVER 200 RANDOMLY GENERATED TEST EPISODES, WITH 95% CONFIDENCE INTERVALS SHOWN TO INDICATE THE RESULTS' CERTAINTY LEVEL. THE BEST RESULT FOR EACH CLASS IS HIGHLIGHTED IN BOLD.

	Train Episodes			**Classification Accuracy**	
Model	Way	Shot	Query	50 epochs	100 epochs
MNs	2	1	5	77.70 ± 5.77%	83.95 ± 5.09%
MNs	2	1	10	79.45 ± 5.60%	85.08 ± 4.94%
MNs	2	1	15	82.32 ± 5.29%	90.13 ± 4.13%
MNs	2	5	5	82.00 ± 5.32%	62.30 ± 6.72%
MNs	2	5	10	**92.85 ± 3.57%**	**90.33 ± 4.10%**
MNs	2	5	15	90.27 ± 4.11%	63.95 ± 6.65%
MNs	2	10	5	61.55 ± 6.74%	64.35 ± 6.64%
MNs	2	10	10	86.83 ± 4.69%	70.83 ± 6.30%
MNs	2	10	15	67.97 ± 6.47%	72.68 ± 6.18%
MNs	3	1	5	84.47 ± 5.02%	88.20 ± 4.47%
MNs	3	1	10	86.60 ± 4.72%	89.82 ± 4.19%
MNs	3	1	15	86.88 ± 4.68%	89.24 ± 4.29%
MNs	3	5	5	54.53 ± 6.90%	56.73 ± 6.87%
MNs	3	5	10	**93.02 ± 3.53%**	52.47 ± 6.92%
MNs	3	5	15	79.76 ± 5.57%	**90.11 ± 4.14%**
MNs	3	10	5	47.10 ± 6.92%	53.87 ± 6.91%
MNs	3	10	10	75.18 ± 5.99%	51.43 ± 6.93%
MNs	3	10	15	52.76 ± 6.92%	62.89 ± 6.70%
MNs	4	1	5	85.17 ± 4.92%	85.55 ± 4.87%
MNs	4	1	10	86.95 ± 4.67%	86.94 ± 4.67%
MNs	4	1	15	**87.08 ± 4.65%**	87.61 ± 4.57%
MNs	4	5	5	46.10 ± 6.91%	**65.48 ± 6.59%**
MNs	4	5	10	31.96 ± 6.46%	31.87 ± 6.46%
MNs	4	5	15	45.02 ± 6.90%	58.25 ± 6.83%
MNs	4	10	5	36.83 ± 6.68%	35.62 ± 6.64%
MNs	4	10	10	35.17 ± 6.62%	35.39 ± 6.63%
MNs	4	10	15	31.67 ± 6.45%	52.98 ± 6.92%