

Received 30 January 2025, accepted 21 February 2025, date of publication 4 March 2025, date of current version 13 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3547573

## RESEARCH ARTICLE

# FiSC: A Novel Approach for Fitzpatrick Scale-Based Skin Analyzer's Image Classification

GUILLERMO CROCKER GARCIA<sup>ID1</sup>, MUHAMMAD NUMAN KHAN<sup>ID1</sup>, AFTAB ALAM<sup>ID1</sup>, JOSUE OBREGON<sup>ID2</sup>, TAMER ABUHMED<sup>ID3</sup>, (Senior Member, IEEE), AND EUI-NAM HUH<sup>ID1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Kyung-Hee University (Global Campus), Yongin 17104, Republic of Korea

<sup>2</sup>Department of Industrial Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea

<sup>3</sup>College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, Republic of Korea

Corresponding author: Eui-Nam Huh (johnhuh@knu.ac.kr)

This work was supported in part by the Ministry of Science and Information and Communications Technology (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2025-RS-2024-00438239; in part by IITP grant funded by Korean Government (MSIT), Global Artificial Intelligence (AI) Frontier Laboratory, under Grant RS-2024-00509257; and in part by Chowis Company Ltd., South Korea.

**ABSTRACT** The Fitzpatrick scale is a widely used tool in dermatology for categorizing skin types based on melanin levels and sensitivity to ultraviolet light. The primary objective of this study is to enhance the accuracy of Fitzpatrick scale classification by addressing limitations in existing methodologies. Current approaches either rely on custom-designed hardware or utilize the Individual Typology Angle (ITA) for image classification; however, these methods typically allow for a one-tone difference in classification and achieve a maximum accuracy of approximately 75%. A primary task for skin tone classification in images, is to apply filters to detect skin regions in an image. However, the filters proposed for detecting skin do not apply to general datasets. In this paper, we propose a novel classification method that employs specialized filters to accurately detect and remove skin surface attributes, such as wrinkles and pores, using a controlled environment dataset obtained from a professional skin analyzer device. Our method involves modeling image features as a nine-dimensional feature vector, followed by a dimensionality reduction process to identify the most influential features and dominant areas within the feature space, enabling deployment on low-power devices. We conducted extensive classification experiments using various Machine Learning algorithms. The results of our cross-validation tests demonstrate a significant improvement in classification accuracy, reaching up to 97%, thereby outperforming state-of-the-art methods without relaxing the accuracy criteria.

**INDEX TERMS** Fitzpatrick scale, skin tone classification, image-based classification, individual typology angle (ITA), feature engineering, skin analyzer device, dermatology image analysis.

## I. INTRODUCTION

Skin tone classification is a vital topic in the cosmetics industry and dermatology [1]. Its importance in medicine is due to its correlation with skin health and skin diseases. In the medical field, relevant studies relate skin tone to malignant melanoma and the risk factors for each tone [2], [3]. Some studies aim to recognize carcinoma as well as melanoma [4]. Many of these diseases can be treated if detected early, as well as other skin diseases are often present differently depending on a patient's skin tone [5], this emphasizes the

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar<sup>ID</sup>.

importance of improving skin tone classification methods. As a numerical classification approach to classify skin types based on melanin and the response to Ultraviolet (UV) radiation, the Fitzpatrick scale [6] is proposed for measuring skin tones. To obtain the Fitzpatrick index, traditionally, an expert examines the skin of the subject and classifies the skin type into one of the six categories described in Table 1.

With the advancement of Artificial Intelligence (AI) and photonics technology, it is now possible to identify skin types without relying on human interpretation. In this context, using Fitzpatrick classification, various solutions have been proposed and could be categorized as device-based [7] and image-based [8], [9].

**TABLE 1.** Fitzpatrick scale categories.

Type	Score	Burn	Tan
I (F1)	0–6	Always	Never (palest; freckles)
II (F2)	7–13	Usually	Minimal
III (F3)	14–20	Sometimes	Uniform
IV (F4)	21–27	Minimal	Always (moderate brown)
V (F5)	28–34	Rarely	Easily (dark brown)
VI (F6)	35–36	Never	Very Easily (deeply pigmented dark brown to darkest brown)

In device-based skin tone classification, hardware systems are used to emit custom-tuned lights, and the properties of the light reflected from the skin are used to determine the Fitzpatrick skin type [7]. On the other hand, image-based skin type classification does not rely on special hardware but instead uses images to differentiate skin types [8], [9]. More recent approaches [10], [11], [12] use the Individual Typology Angle (ITA) [13], which is a rule-based classifier, for Fitzpatrick scale measurement. Utilizing Fitzpatrick Scale-based skin analysis offers various applications for smart devices. It provides personalized skincare recommendations, tracks sun exposure, and suggests beauty products based on skin type. Additionally, it assists in health monitoring by detecting skin tone changes, enables teledermatology consultations, and enhances virtual try-ons in fashion and e-commerce. These features make the skin care devices a versatile tool in both personal care and healthcare.

One of the most challenging issues in the Fitzpatrick classification problem is the measurement accuracy [3], for which device-based approaches score better than image-based approaches due to the inconsistency in image qualities and the high variability of how the image is captured [14], [15], [16]. For improving the Fitzpatrick measurement accuracy, device-based approaches [7] use a controlled environment covering the skin and use specialized lights for their analysis. However, the results of these devices are measured during sampling time, and moreover, these devices do not use images for analysis because they are based on spectrophotometry [2], [7], [17] or photo-acoustic [18], [19], [20], [21] technologies.

To solve these problems, we propose a novel method for Fitzpatrick classification. We use a professional skin analyzer device that can provide images to be analyzed. Using labeled images from a controlled environment, we design novel filters to detect skin surfaces in images. One of the challenging issues is the fact that the optical zoom is high in device-taken images, so the skin texture (such as wrinkles or pores) are captured in a darker tone and it can result in the image being misclassified. After the filtering phase, we propose a novel feature engineering algorithm to identify influencing features and then perform the Fitzpatrick scale classification utilizing Machine Learning (ML) algorithms. Furthermore, the proposed method is designed to make the Fitzpatrick classification more accurate, and our methodology can be extended to other specialized skin analyzers that are able to produce images in a controlled environment. We also present a unique reflection reduction method using a novel

discriminating criteria. During experiments, first, we used feature engineering to extract and select skin-related features from the color spaces suggested in [22]. However, after a deep study and analysis, we identified the most influential features through a correlation study, supported by Principal Component Analysis (PCA). Finally, we conducted extensive experiments on the influencing features in the reduced dimension.

In theory, our proposed methodology is similar to other works, such as [11], [12] in the sense of using a controlled environment and a unified value per image. Yet, in our pursuit to enhance accuracy and ensure accessibility on resource-constrained devices, we encountered numerous challenges necessitating innovative approaches and simplifications to effectively address the Fitzpatrick classification problem. Specifically, our focus lies in overcoming four key challenges, outlined as follows:

#### A. CHALLENGE 1

The primary challenge lies in the inconsistency of image qualities and the wide variability in how skin images are captured. This variability significantly hampers the process of identifying suitable features for classification due to the inherent nature of skin images.

#### B. CHALLENGE 2

The second challenge entails optimizing the proposed methodology to be efficient and lightweight, ensuring it can perform classification effectively on devices with limited resources. Considerations for real-time processing and energy efficiency are also crucial for deployment in resource-constrained environments.

#### C. CHALLENGE 3

Accurately detecting the skin surface while avoiding interference from skin textures and light reflections poses a significant challenge in preventing misdetections. Robust techniques for distinguishing genuine skin features from artifacts are essential, requiring sophisticated techniques capable of discerning subtle differences in texture and illumination.

#### D. CHALLENGE 4

Lastly, improving the overall accuracy of classification is paramount in skin classification, especially considering the medical implications involved in the procedure.

To meet our objectives, particularly in addressing the aforementioned challenges, we have devised a novel methodology for Fitzpatrick scale-based image classification employing ML techniques, specifically tailored for low-powered skin analyzer devices operating in controlled environments. The proposed methodology aims to improve measurement accuracy, detect and remove skin surface attributes such as wrinkles and pores, perform an effective dimensionality reduction by identifying the most influential features for

accurate classification, and create a unified 3D vector that represents each image to make it lightweight enough to deploy on devices with limited resources, as well as reduce the overall ML techniques complexity by reducing its parameters.

Our main contributions are listed below:

- We propose a novel and automatic Fitzpatrick scale-based image classification technique for low-powered skin analyzer devices that work in controlled environments.
- We designed a novel skin filter capable of identifying skin surface and textural imperfections that can result in a misclassification to a darker skin tone.
- We propose a novel reflection removal technique to avoid the reflections caused by skin sebum in images that can bias the classification to a brighter skin tone by using a novel discriminating criteria for detecting reflections.
- With the aim of performing dimensionality reduction and using feature engineering to improve the performance by reducing classifier parameters, we identify the three most influencing features that can show similar accuracy to the feature vector containing nine selected features from image channels.
- We evaluated our proposed system through extensive experiments, classifying the Fitzpatrick scale using various ML classification methods.

## II. RELATED WORK

Dermatologists utilize the Fitzpatrick scale [6], which rates skin tone from I (F1)-lightest- to VI (F6)-darkest-, to categorize susceptibility to sun exposure. Of course, there are differences in skin tones between Caucasians and African Americans. In addition, independent of variations in skin tone, there are differences in facial morphology by race, gender, and person.

There are two main streams among methods for classifying the Fitzpatrick scale: device-based and image-driven. In the device-based stream, spectrophotometry is the standard technique for measuring skin tone [2]. Commercial devices, such as the DSM II ColorMeter (Cortex Technology ApS, 9560 Hadsund, Denmark) [23], measure melanin levels using the method proposed by [17], which relies on red, green, and blue wavelength remittances. The accuracy of an Skin Tone Meter (STM) prototype device that emits light at a wavelength of 460 nm, where melanin absorption dominates over other skin chromophores such as oxyhemoglobin and water, was examined in [7]. By measuring the intensity of reflected light, the STM device calculates the subject's skin Pigmentation Index (PI) which is then mapped to one of the skin types in the Fitzpatrick scale. Although the results show promising improvements against the conventional Fitzpatrick questionnaire method, however, mapping of PI indices to Fitzpatrick skin types is not scientifically proven and, in this case only applicable to the chosen STM device. Other

approaches are based on photoacoustic and reflectance over skin [18], [19], [20], [21].

On the other hand, the rapid advancement of computer vision and AI has enabled the classification of the Fitzpatrick skin type, leading to the image-driven approach. However, one big challenge for accurate measurement of skin type using images is varying illuminations. To mitigate the effect of different illumination conditions, [8] proposed a method to extract skin tone from skin images of COVID-19 patients by normalizing the lightness component in the hue-saturation-lightness scale. Although lightness normalization allows an approximation of the true skin color, however, the accuracy of their approach is debatable as only one normal skin pixel is selected from each patient image. On the contrary, instead of using a single pixel, [9] identifies their Region of Interest (ROI) from full-face images using face landmark detection. Specifically, three regions are selected from the forehead, left cheek, and right cheek. To reduce the impact of changing illuminations and backgrounds, [9] decomposes reflected light into specular and diffuse components where the latter is used to construct a skin color metric. Their approach shows greatly reduced intra-subject variation in skin type estimation even if the illumination and background vary drastically. Nevertheless, due to a lack of ground truths, the accuracy of skin type classification against the true skin color of the subjects is unknown in their work as well.

Utilizing color information and making use of the fact that skin tone colors share some common characteristics [22] which are also useful for the Fitzpatrick classification. However, before skin tone classification, the problem of detecting skin area appears. The skin detection problem is typically solved by implementing computer vision-based filters and setting an ROI as [24], [25], and [26], however, these techniques just remove noise from the skin and do not determine the skin tone. Other studies such as the LBP-based methods [27], [28], [29] aim to classify general texture features e.g. fabric types. However, skin textures differ from general textures depending on many factors like the part of the body where the image is taken, the age of the subject, and skin condition. Moreover, these techniques do not apply to our work since the aim of our study is to classify skin tone and not to classify the general texture labels or in our case the skin texture labels. Certain skin identification techniques can be used with grayscale images, such as [30], but they only produce decent results when used with high-resolution images because they study local texture properties. The appropriate color space transformation can be used to create rules that discriminate between skin pixels and other kinds of items and then the skin tone is classified using ITA [10], [12].

These methods necessitate the use of RGB color images because they are based on color information. This family of methods could not be used in a particular situation where other sensors are added since it was not possible to build a way that could function with the grayscale camera output

from the Single-photon Avalanche Diode (SPAD). There is also an important difference between cameras and settings that makes difficult the discrimination between skin and noise.

Other approaches aim for face segmentation, also known as face parsing, which is the process of tightly segmenting an input image of a face into regions that correspond to the various parts of the face and the backdrop [11], [12], [31]. This is done densely by labeling the pixels or assigning a label to each pixel. We are not addressing this issue since our dataset is collected from a specialized device, i.e., Dermo Prime (DP) [32] that can take images from any part of the body as well as the face. Groh et al. [10] annotated dermatology clinical images with Fitzpatrick skin type labels. For their experiments and testing, they used a filtered and masked image approach, applying the overall non-masked pixels' ITA [13] for Fitzpatrick classification. This method involves a CIELAB transformation [33] to compute the output using a specific formula, followed by ITA, a decision rule-based approach that classifies skin type based on predefined ranges. They observed that the skin type in the images used to train the model had an impact on the accuracy scores across Fitzpatrick skin types. They also constructed a deep neural network model to detect skin disorders. References [11] and [12] proposed a method for classifying the Fitzpatrick scale over face images. They used color transformation and filters as preprocessing ITA for classification. They used the mean value of the skin pixels and then compute ITA.

Existing literature does not address issues like specialized filters for low-power devices. Moreover, their filtering processes are expensive in terms of computation and can slow down the performance of the overall analysis. On the other hand, the images acquired are from uncontrolled environments which makes it hard to use existing classification techniques for the Fitzpatrick scale measurement. To solve these problems, we propose a novel method for Fitzpatrick classification. The images are generated by a professional skin device for analysis. We propose filters to identify skin surfaces in images by using labeled images generated in a controlled environment. We also address the issue of feature engineering to find influencing features and then perform Fitzpatrick scale-based classification using ML algorithms.

**TABLE 2.** CDP analyzer specifications.

Camera Resolution	1 MP
Magnification	30x
Image Size	640×480
Image Format	JPEG
Imaging modes	PPL, XPL, UVL
Working Distance (camera lens to skin surface)	1.7 cm
Field of View	2.3cm in diameter
Skin sample size	2.3cm in diameter

### III. METHODOLOGY

The abstract view of the proposed Fitzpatrick Scale based Classification utilizing ML techniques (FiSC) framework for

skin tone classification is shown in Fig. 1 and comprises four phases i.e., Data Preparation, Data Pre-processing, Feature Engineering, and Training and Applying Classification Models.

#### A. DATA PREPARATION

The proposed techniques are best suited to classify the images being produced by skin analyzer devices, e.g., CDP [32]. Professional skin tone dataset images are captured under a custom-tuned Plain Polarized Light that best preserves the colors of the true skin tone. Fig. 2 shows the wavelength emitted by the skin analyzer device when taking skin images with Plain Polarized Light, which is simply defined as light visible to the naked eye. Table 2 shows the specifications of the device.

**TABLE 3.** Number of images in each category of the FiSC dataset and FiSC aug-dataset.

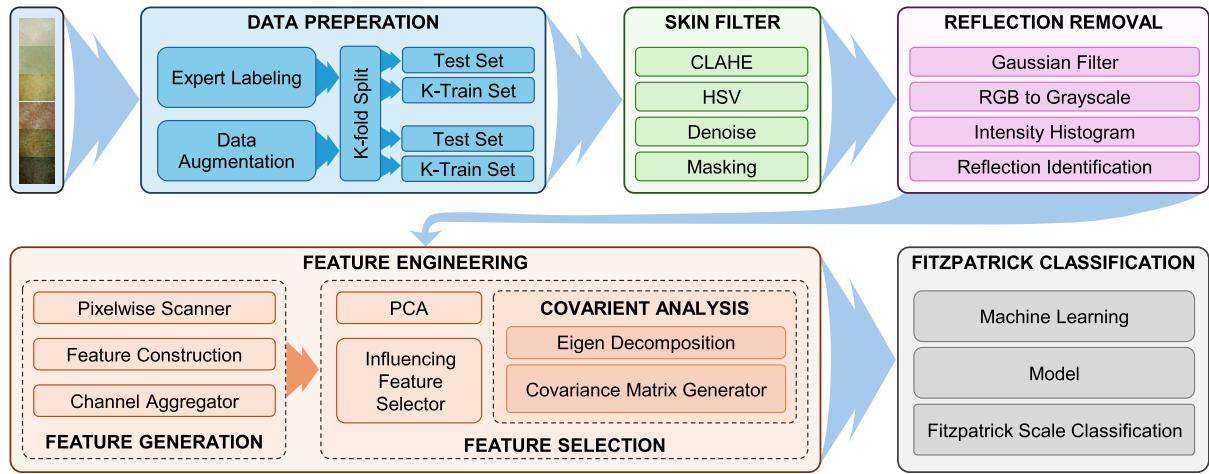
Category	FiSc Dataset	FiSc Aug-Dataset
F1	1507	1507
F2	619	1500
F3	307	1500
F4	143	1500
F5	24	1500
F6	137	1500
<b>Total</b>	<b>2800</b>	<b>9007</b>

Using the skin analyzer device [32], we produced an image dataset called FiSC Dataset consisting of 2800 images. The images were human-labeled according to Fitzpatrick scale standards and validated by domain experts. Based on Scale AI's dynamic consensus approach, two annotators labeled the Fitzpatrick skin type of each image, with the help of this device that counts with an optical zoom of 20X. Each image is evaluated by annotators more objectively among ethnic groups [15]. Sample labeled images of the dataset are shown in Fig. 3. The dataset includes a diverse range of images collected from candidates spanning multiple ethnic groups including Eastern Asians, Southeastern Asians, Caucasians, Hispanics, and South Africans using the Skin Analyzer device.

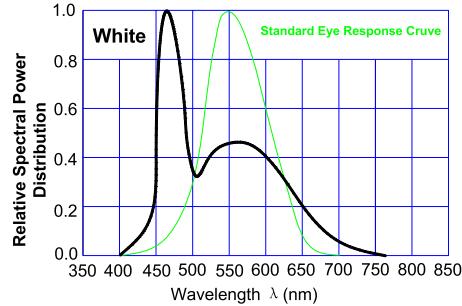
We use data augmentation techniques to balance the categories since there are a few images in Fitzpatrick prototypes F5 and F6, thus, we round the number of samples to 1500 samples per prototype. The augmentation techniques used are: flip top bottom, flip left-right, and rotate 180°, we did not use scaling since the device already has an optic zoom of 20x and this can distort the image quality. We call the augmented dataset as FiSC-Aug Dataset. The labeled images for each prototype category are shown in Table 3.

#### B. DATA PRE-PROCESSING

Since the images in the dataset are taken using a specialized device, but the optical zoom is high, the skin texture, like wrinkles or pores that are captured in a darker tone can result in misclassification of the real skin tone. To alleviate this issue, we propose a skin filter based on image transformations and reflection removal technique.



**FIGURE 1.** Block diagram of the proposed architecture.



**FIGURE 2.** Wave length emitted by the CDP analyzer.

### 1) SKIN FILTER

Finding the skin textures clearly is imperative for skin surface detection. To achieve this, we tested the techniques presented by [24] and [26] on the FiSC dataset. Since the FiSC dataset uses images taken from optical zoom-based devices in a controlled environment, the results of using the techniques presented in [24], [25], and [26], were not satisfactory, since these techniques are oriented to locate the skin as ROI and mask otherwise. However, we are using masking techniques to differentiate skin and skin textures. In this regard, we propose a novel approach that best suits datasets being produced by optical zoom-based devices. We utilize an Adaptive Histogram Equalisation (AHE) technique [34] to detect the skin surface. More specifically, we use the Contrast Limited Adaptive Histogram Equalization (CLAHE) method proposed by Zuiderveld and Karel [35], which is a type of AHE in which the amount of contrast amplification is constrained, thereby limiting the range of contrast. After the skin surface detection, the resulting image from CLAHE is converted to Hue, Saturation, and Value (HSV) color space. For skin texture and surface, we exploit the notion of depth proposed by [36] for stereoscopic images. This can be explained as the difference in HSV color space that is related to the depth that can be used in a single camera with a fixed

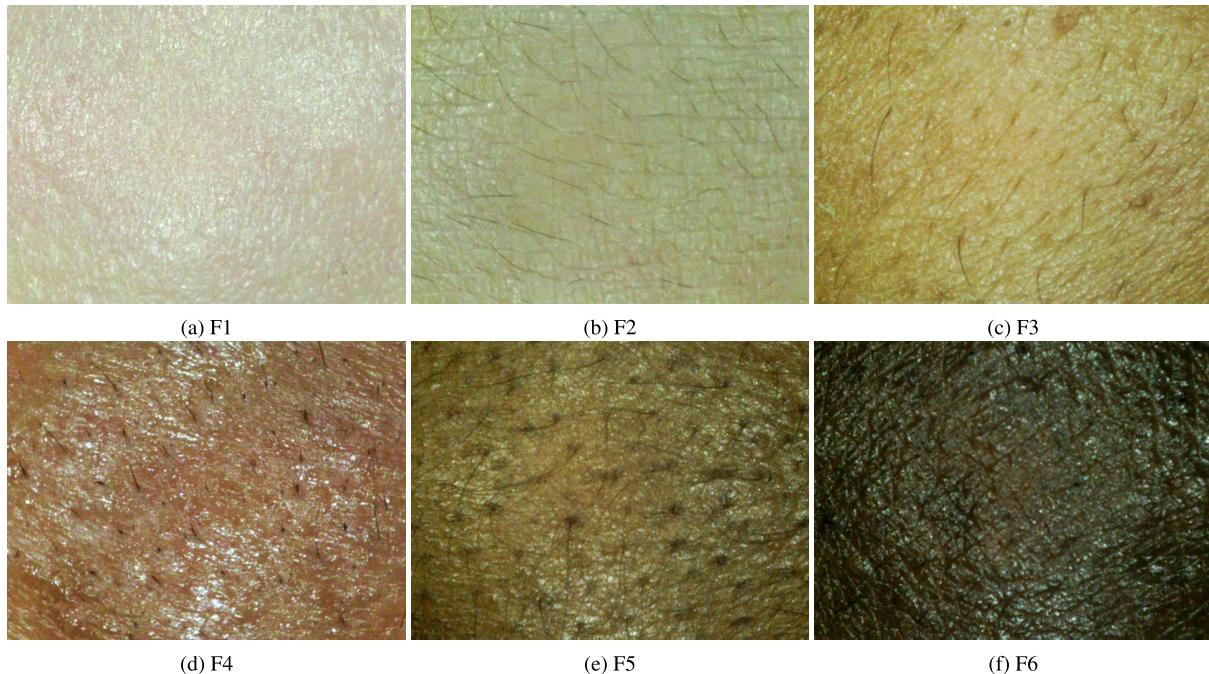
distance from the lens to the skin, this difference appears in skin marks and not in flat skin. Having the HSV image derived from the CLAHE, this image can contain some noise. We denoise the image through a binary threshold that we call  $R''G''B''$  (denoised) image. In the  $R''G''B''$  image, the patterns of texture are clearly visible in green and black as well as the skin surface in red as shown in Fig. 4. Finally, we mask the textures and overlap this mask into the original image, which is then fed to the reflection removal step. We set the filtering condition as a function  $f_c$  in (1), where if the conditions are fulfilled, the value is true, and if not, the value is false.

$$f_c = \begin{cases} RGB_{ijk}, & \text{if } R''_{ij0} = 255 \wedge G''_{ij1} = B''_{ij2} = 0 \\ [0, 0, 0], & \text{otherwise} \end{cases} \quad (1)$$

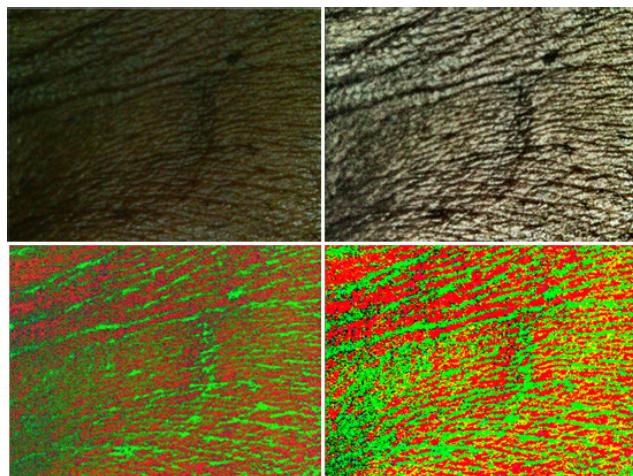
Using the matrix representation of the channels in color spaces and normalizing the original image, the RGB channels represent the original image. The  $R'G'B'$  are the  $R''G''B''$  channels, which are converted to RGB color space. We assign the resultant values to the resultant matrix  $M_{ijk}$ . The filtering process is described in (2), where the original image has dimensions  $i \times j, k = [0, 1, 2]$ . If the value is false for  $M_{ijk}$  then assigned value is  $[0, 0, 0]$ , otherwise the assigned value is  $[R_{ij0}, G_{ij1}, B_{ij2}]$  that is the value of the original image.

$$\begin{cases} R''_{ij0} G''_{ij1} B''_{ij2} \\ R_{ij0} G_{ij1} B_{ij2} \end{cases} | f_c \rightarrow \begin{cases} M_{ij0} \\ M_{ij1} \\ M_{ij2} \end{cases} \quad (2)$$

Algorithm 1 illustrates the proposed skin filter process. First, the CLAHE technique is applied to the original RGB image, resulting in what we call the CLAHE image. Then, the CLAHE is converted to HSV CS, resulting in HSV. The HSV image is converted to RGB CS channels, resulting in what we call the  $R'G'B'$  image. To remove the noise, a binary threshold is applied to the  $R'G'B'$  image, resulting in what we call  $R''G''B''$ . Finally, it is necessary to mask the image as described in  $f_c$ , resulting in the masked image  $M$ . The returned



**FIGURE 3.** Fitzpatrick image sampling from device (a) photo-type F1 (b) photo-type F2 (c) photo-type F3 (d) photo-type F4 (e) photo-type F5 (f) photo-type F6.



**FIGURE 4.** Transformation process for identifying skin and textures: (Top-left) original skin image, (Top-right) CLAHE transformation, (Bottom-left) HSV transformation, (Bottom-right) R'G'B'-D denoised.

image is the masked image from the Skin filter's returned image with the generated mask overlapped.

## 2) REFLECTION REMOVAL

After all skin filters have been applied, we propose a method for removing reflections as many images have them as a result of the device's specialized lights. We use a Gaussian filter to smooth and reduce the noise in the image. To deploy a Gaussian Filter to an image, we define the size of the kernel which is a matrix built up through the values computed by the Gaussian function in (3), and is used for convolving all over

---

### Algorithm 1 Skin Filter

---

**Input:** RGB  
**Output:** Masked Image

- 1 CLAHE  $\leftarrow$  Apply to CLAHE RGB
- 2 HSV  $\leftarrow$  Convert CLAHE
- 3 R'G'B'  $\leftarrow$  Convert HSV to RGB
- 4 R''G''B''  $\leftarrow$  Threshold R'G'B'
- 5 M  $\leftarrow$  Apply  $f_c$  to R''G''B''

---

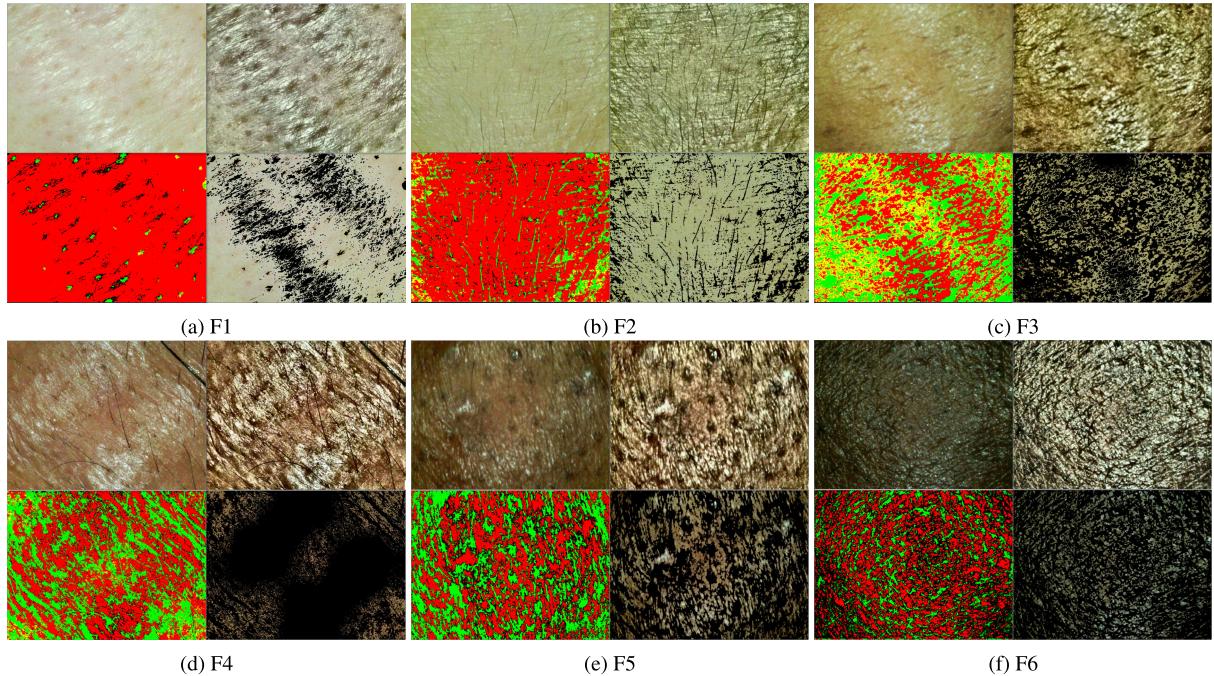
the image.

$$h(l, k) = \frac{1}{2\pi\sigma^2} e^{-\frac{l^2+k^2}{2\sigma^2}} \quad (3)$$

Here,  $l$  is X coordinate value,  $k$  is Y coordinate value of the Gaussian kernel and  $\sigma$  is the Standard Deviation. We apply the Gaussian Filter using the Gaussian kernel  $h(k, l)$  in (3). The Gaussian Filter is computed as a weighted summation of the input pixel values  $f(i + k, j + l)$ , and is applied to a resulting image  $g(i, j)$  in (4).

$$g(i, j) = \sum_{k,l} f(i + k, j + l) h(k, l) \quad (4)$$

To calculate the range of pixels with brighter values that can be perceived as reflection, we compute a mono-channel-based intensity histogram over the grayscale image after the RGB to grayscale conversion. Then, using the histogram to get the highest hit bucket value, we scan the grayscale image. In the context of image processing, the word "intensity histogram" refers to a histogram of the pixel intensity



**FIGURE 5.** Filter behavior per photo-type (a) photo-type F1 (b) photo-type F2 (c) photo-type F3 (d) photo-type F4 (e) photo-type F5 (f) photo-type F6.

(grayscale) values calculated from an image. This histogram is a graph that shows the number of pixels in the input picture at each distinct intensity value. We define a composite intensity histogram as  $H$ , and thus as a set, we define the composite intensity histogram of an image as  $C$  in (5), where  $h(i)$  is the histogram value at intensity  $i$  (the number of picture pixels  $n(i)$  bearing the intensity grey level value  $i$ ), and  $\ell$  is the total number of different intensity levels called buckets that an image can have for the intensity quantization. We then define  $C(k)$  in (6).

$$C = \{h(i) \mid i = 0, 1, 2, \dots, \ell - 1\} \quad (5)$$

$$C(k) = \sum_{i=0}^k h(i) \quad (6)$$

This model can be used to represent the histogram if we define it as a probability distribution. Let  $\psi$  be an image with an integer pixel intensity range of 0 to  $\ell - 1$ , represented by a  $m_r \times m_c$  matrix, where  $\ell$  is the total number of possible intensity values (typically between 0 and 256). Then  $p$  represents the normalized histogram of  $f$  with a bin for each possible intensity as shown in (7).

$$p_n = \frac{\psi}{N} \quad (7)$$

Here,  $\psi = \{0, 1, \dots, \ell - 1\}$  is the total pixel number with intensity  $n$  and  $N$  is the total number of pixels. Using (7), the probability distribution can be transformed from a discrete variable to a continuous variable. The motivation for this transformation comes from considering the intensities of  $f$  and  $g$  as continuous random variables  $X, Y$  on  $[0, \ell - 1]$  with

$Y$  defined in (8).

$$Y = T(X) = \int_0^X (\ell - 1)p_X(x)dx \quad (8)$$

---

#### Algorithm 2 Reflection Removal

---

**Input:** Masked image from Filter  
**Output:** Masked Image

```

1 G ← GaussianFilter(G)
2 G ← RGB2Grayscale           // Mono Channel
3 H ← IntensityHistogram(G)   // Avoid 0 (black)
   value
4 i ← MaxHitBucketValue(H)
   // Discriminate reflection pixels
5 foreach pixel δ ∈ G(x, y) do
6   | if δ > (i × C) then
7   |   | RGB(x,y) ← 0,0,0          // black
8   | end if
9 end foreach

```

---

We use this continuous representation to define the maximum hit bucket value  $i$  from the histogram, which is the local maximum of the probability distribution in the continuous variables where the derivative slope is 0, and we define this in (9).

$$Y = i = T(X) = \frac{d}{dx} p(x) = 0 \quad (9)$$

After the intensity histogram is applied, it is used to discriminate the reflections in the following way: Let  $\delta$  be

individual pixels  $\in G(x, y)$ , the discrimination criteria is defined in (10).

$$\delta > (i \times C) \quad (10)$$

Here,  $i$  is the maximum hit bucket value from the intensity histogram, and the factor is a constant  $C$  that can be obtained using the local maxima. The Skin-Fine filter's returned picture, with the generated mask overlapped, is the masked version of that image. The behavior of our filtering method is shown in Fig. 5. The proposed technique of Reflection Removal is shown in Algorithm 2. The Gaussian filter in (4) is applied and then the resultant image is converted to grayscale. The intensity histogram in (6) is created and by applying the criteria, the max hit bucket is identified as in (9). Finally, the pixel values are being discriminated as defined in (10). If the criteria is not accomplished, it is set as black, which means that it is detected as reflection.

**TABLE 4. Feature description per channel.**

Feature	Description
r	red channel unified value
g	green channel unified value
b	blue channel unified value
L	lightness ( $L^*$ ) channel unified value
A	$a^*$ channel unified value
B	$b^*$ channel unified value
H	Hue channel unified value
S	Saturation channel unified value
V	Value channel unified value

### C. FEATURE EXTRACTION

Our goal is to estimate skin tone through correlation with the Fitzpatrick labels, and for this purpose, ITA is widely used. ITA is a measure of the topology angle calculated from statistical features of pixels in an image. Previous studies by [10] and [12] used ITA to evaluate masked images over the non-masked areas. In our proposed method, similarly, we filter the image to detect skin pixels and mask the image otherwise. We extract these skin pixels to perform feature engineering. In this context, we propose color space transformations and find the most relevant features. However, using individual pixel values for feature engineering is not practical due to the large number of pixels per image. To address this, we suggest using a representative value such as the average, which is feasible in our proposed method since most filtered pixels have similar values.

#### 1) FEATURE EXTRACTION FOR DETECTED SKIN

The masked image has been transformed in the way that it is identified as the skin on images and is set as a black mask on everything else. Over these RGB Color Space (CS) images, two specific transformations are performed: HSV and CIELAB (LAB). Each image (and channel) undergoes a complete pixel-wise scan to extract a unified value for each channel. Note that there are nine channels total per image after the CS transformations. For simplicity, we will refer to each channel's unified value as shown in Table 4. This image

only contains what is considered as skin, everywhere else is the black mask. To extract the unified value, we noted after experiments that the most representative value for unified value calculation is the average value. This extracted unified value per channel from the input image can be modeled as a 9D space feature vector as follows  $[r, g, b, L, A, B, H, S, V]$ .

#### Algorithm 3 Feature Extraction

---

**Input:** Channel from Color Space  
**Output:** Unified value per channel

```

1 C ← Channel                                // Mono Channel
2 counter ← 0
3 values ← 0
4 foreach pixels of image p ∈ C(x, y) do
5   if p > 0 then // 0 is pixel of mask
6     Increase counter
7     values ← values + p
8   end if
9 end foreach
10 if counter > 0 then
11   Unified value ← values/counter
12 end if

```

---

Algorithm 3 shows the step-by-step process of feature extraction. First, we have each channel per CS and set a *counter* and *values* where we will sum up the individual pixel values. Then, we check that there are skin pixels that after filtering and reflection removal are not masked (pixel value is  $> 0$ ), and increase the counter as well as sum up the values. Finally, we check if there is not a “black” image where there is not any skin detected and return the unified (average) value.

#### 2) INFLUENTIAL FEATURES SELECTION

Having the 9D feature vector, the feature vector is analyzed using PCA for checking covariance among the features and checking for influencing features that have no correlation for performing a dimensionality reduction on the most influencing features. These influencing features are located in clusters that we call the dominant areas. Our dataset is modeled as 9D as described in (11).

$$x^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} = \begin{bmatrix} r \\ g \\ b \\ L \\ A \\ B \\ H \\ S \\ V \end{bmatrix} \quad (11)$$

To reduce the dimensionality, our feature engineering approach should find the most influencing features. With this aim, first, we applied a covariance study among the 9D features over the raw dataset images to detect potential

influencing features. The covariance between two elements is determined using (12).

$$\sigma_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (12)$$

Using this covariance calculation, the covariance matrix  $\Sigma$  is defined as a  $d \times d$  matrix in which each element indicates the covariance between two features. This covariance matrix can be calculated using the (13), where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  is the mean vector.

$$\Sigma = \frac{1}{n-1} ((X - \bar{x})^T (X - \bar{x})) \quad (13)$$

The eigenvectors  $v$  are represented as  $n \times 1$  matrices, and the linear transformation can be represented in the form of an  $n \times n$  matrix, and having the eigenvalue in (14) for the linear transformation stated above can be expressed as the matrix multiplication as defined in (15).

$$T(v) = \lambda v \quad (14)$$

$$Av = \lambda v \quad (15)$$

Then, having eigenvectors  $v$  and eigenvalues  $\lambda$ , the matrix  $A$  can be decomposed using them, which is called eigendecomposition, and the correlation among features is calculated. These eigenvectors and eigenvalues that are contained in the covariance matrix represent the fundamental elements of a PCA analysis in the following way: The directions of the newly estimated feature space are determined by the eigenvectors, which are principal components defined in PCA, and by the eigenvalues, which determine the magnitude of these main components. In other words, the variance of the data can be explained by the eigenvalues along the new feature axes.

PCA aims to perform this eigendecomposition and maximize the variance and it is calculated with (13). Each value in the mean vector, an  $n$ -dimensional vector, reflects the sample mean of a feature column in the dataset. In the higher-dimensional space, our goal is to reduce the number of features for the classification, which is done through our feature engineering approach. We select the most relevant skin features which we call influencing features of each CS to be near the axes of the PCA. These influencing features are then clustered in dominant areas.

#### D. CLASSIFICATION MODELS

ML is the field that studies algorithms that learn from data. The goal is to construct a function  $f(x)$  that can map the input feature  $x$  to the response variable  $y$ . When  $y$  is a category, the ML task is called classification. In this study, we used ML algorithms to classify the Fitzpatrick scale type. We use two sets of features as input for evaluating our proposed method. The input  $x$  for the ML algorithms is the extracted feature vector representing the unified value per channel. In this work, we use seven ML techniques, i.e., Logistic Regression (LR) [37],  $k$ -Nearest Neighbours (KNN) [38], Decision Trees

(DT) [39] and Support Vector Machines (SVM) [40], two ensemble algorithms: Random Forest (RF) [41] and Gradient Boosting Machines (GBM) [42], and Dense Neural Network (DNN) [43].

##### 1) LOGISTIC REGRESSION

The goal of logistic regression is to classify the response variable into one of six skin types [37]. Because we have more than two classes, this method is called multinomial logistic regression. For a skin type  $c \in C$ , we model the probability of a sample belonging to that class using (16).

$$P(Y = c | X = x) = \frac{e^{\beta_c^T x}}{1 + \sum_{l \in C} e^{\beta_l^T x}} \quad (16)$$

where the vector  $\beta_c$  represents the regression coefficients for each class  $c$ . The logistic regression function classifies a sample into the skin type with the highest estimated probability.

##### 2) K-NEAREST NEIGHBORS CLASSIFICATION

The KNN classification algorithm is a pattern recognition technique that seeks to identify the  $k$  closest relatives in subsequent cases using training datasets [38]. When KNN is used in classification, it estimates the  $k$  closest data points that are near the sample  $x$ , represented by  $\eta$ . Then, it estimates the conditional probability in (17) for class  $k$  as the fraction of neighboring points whose class is equal to  $k$ .

$$P(Y = c | X = x) = \frac{1}{k} \sum_{i \in \eta} I(y_i = c) \quad (17)$$

where  $I$  represents the indicator function, which returns 1 if the condition inside the function evaluates to true, and 0 otherwise.

##### 3) SUPPORT VECTOR MACHINES CLASSIFICATION

The Support Vector Machine (SVM) classification seeks to train and categorize data in accordance with polarity levels [40]. The goal is to find the hyperplane that can accurately divide the data. The optimum hyperplane is thought to be the one with the greatest distance between each class. In order to optimize the machine learning rate, the hyperplane is formulated in (18), where  $x$  is an array of vectors and  $w$  is a constants array.

$$w \times x + b = 0 \quad (18)$$

Once the hyperplanes are established, the following definition of the hypothesis function for making predictions are available as (19), where the classifier will assign the point above or on the hyperplane to a class +1 and the point below to a class -1, respectively.

$$h(x_i) = \begin{cases} 1, & \text{if } w \times x + b \geq 0 \\ -1, & \text{otherwise} \end{cases} \quad (19)$$

This method can adapt many hyperplanes; due to the different features and the variety of the images, this is a

complex dataset, and many hyperplanes should be required, making this a huge classification model in terms of size and complexity to compute as well. When the dimensionality was reduced, the model showed better performance, as in our influencing feature selection approach.

#### 4) DECISION TREE CLASSIFICATION

Decision trees are a supervised learning technique that aims to partition the feature space into more homogeneous regions with respect to the response variable  $y$ . The impurity of feature space regions is measured with different measures, but the most common is the Gini index [39]. Decision trees emulate the human decision process and are organized in a hierarchical structure with a root node that moves to branches that represent conditional evaluations using features, and then it moves to the leaves that are used to define boxes or regions  $R_m$  with  $m \in M$  representing each leaf node in the tree. Decision trees predict the skin type of each instance to the most commonly occurring class in the enclosed region  $R$ . It estimates the conditional probability of class  $c$  in node  $m$  as in (20), where  $N_m$  represents the number of training samples that fall in the region  $R_m$ .

$$P(Y = c | X = x) = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = c) \quad (20)$$

#### 5) RANDOM FOREST CLASSIFICATION

Ensemble learning is a set of algorithms that combine the results of different single predictive base models for improving classification performance. Bagging trees (Bootstrap aggregating) train several decision trees in parallel and combine their predictions by majority voting. Random Forest are improved bagging ensemble models that de-correlate their internal decision trees by randomly selecting a subset of features for training each base classifier [41]. This improves the generalization performance of the final models as well as reduces the variance, making Random Forests one of the most powerful and widely used ML algorithms in practice. Random Forests built  $B$  decision tree classifiers  $h(x)$  and used a majority vote scheme for computing the conditional probability of a new instance belonging to class  $c$  as in (21).

$$P(Y = c | X = x) = \frac{1}{B} \sum_{b \in B} h(x) \quad (21)$$

#### 6) GRADIENT BOOSTING CLASSIFICATION

Gradient Boosting Machines (GBM) [42] is a different flavor of ensemble learning. GBMs are built sequentially, and each newly trained base classifier attempts to compensate for the errors of the previously trained models. Boosting is equivalent to performing a forward stage-wise additive expansion using an exponential loss function. Thus, GBM can use an arbitrary differentiable loss function. In a similar way to logistic regression, GBM predictions work to model the probability of an instance belonging to class  $c$  using (16). The strategy used is a greedy stage-wise approach, in which

at step  $b$  there is an imperfect model with the form in (22). This way, GBM improves the function  $H_n(x)$  by fitting a new model  $h_n(x)$  over the so-called pseudo residuals.

$$H_n(x) = H_{n-1}(x) + w \times h_n(x) \quad (22)$$

#### 7) DENSE NEURAL NETWORK CLASSIFICATION

Dense Neural Network (DNN) is an Artificial Neural Network technique that uses backpropagation and multiple layers between the input and output layers [43]. To perform similarly to the human brain, in a DNN the following elements are defined: neurons, synapses, weights, biases, and functions. With these elements, DNN tries to make a linear relation among the neurons where prediction is based on selecting the highest probability for the next layer. This relation can be represented in (23), where  $w$  is the weight matrix and  $b$  is the bias.

$$y = w \times x + b \quad (23)$$

For training our data, we designed a model with 32 input neurons and 3 dense layers, with the last output layer having 6 neurons which is the total number of classes. We used ReLU as an activation function. We regularised the neuron outputs among the layers using Adam optimizer, and for the loss function, we used the Sparse Categorical Cross-entropy function, which is a setting widely used for classification.

## IV. RESULTS AND DISCUSSION

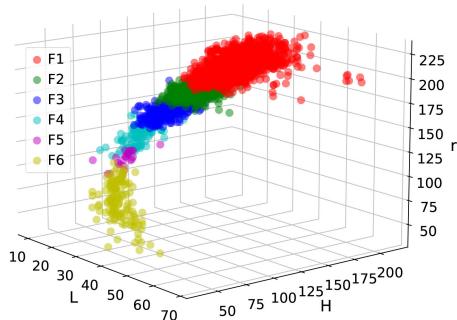
### A. DATASET

While using skin analyzer device [32], we produced an image dataset called FiSC Dataset consisting of 2800 images. The images were human-labeled according to Fitzpatrick scale standards and validated by domain experts. Based on Scale AI's dynamic consensus approach, two annotators label the Fitzpatrick skin type of each image. Some sample-labeled images of the dataset are shown in Fig. 3. The dataset includes a diverse range of images collected from candidates spanning multiple ethnic groups including Eastern Asians, Southeastern Asians, Caucasians, Hispanics, and South Africans using the Skin Analyzer device.

The labeled images for each phototype category were distributed in the following way: F1 – 1507 samples, F2 – 619 samples, F3 – 307 samples, F4 – 143 samples, F5 – 24 samples, and F6 – 137 samples, as shown in Table 3. We used data augmentation techniques to balance the categories since there are a few images in Fitzpatrick prototypes F5 and F6, thus, we round the number of samples to 1500 samples per phototype. The augmentation techniques used are: flip top bottom, flip left-right, and rotate 180°. We did not use zoom since the device already has an optic zoom of 20X and this can distort the image quality. In this study, the augmented dataset is called FiSC Aug-Dataset.

### B. EXPERIMENTAL SETTING

All of the experiments were carried out on a machine with an Intel(R) Core(TM) i7 processor, 16.0 GB of RAM, and



**FIGURE 6.** rLH extracted from dataset, plot showing the stratum for each class.

**TABLE 5.** rLH accuracy/k fold = 5 on FiSC dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic Regression	80.21	0.0126	87.48	0.0075	87.64	0.0041
K-nearest Neighbors	80.88	0.0202	86.58	0.0136	87.03	0.0117
(Kernel) SVM	82.28	0.0144	89.22	0.0034	90.24	0.0058
Decision Tree	79.33	0.0226	86.50	0.0083	87.60	0.0079
Random Forest	80.66	0.0218	88.10	0.0092	89.03	0.0090
Gradient Boosting	81.08	0.0158	88.82	0.0101	90.15	0.0129
DNN	81.55	0.0135	90.01	0.0100	<b>91.64</b>	0.0104

**TABLE 6.** rLH accuracy /k fold = 5 on FiSC-aug dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic regression	71.38	0.0070	81.81	0.0054	86.11	0.0055
K-nearest neighbors	87.41	0.0085	89.73	0.0075	93.65	0.0064
(Kernel) SVM	76.00	0.0105	84.73	0.0050	89.31	0.0060
Decision tree	67.30	0.0084	80.20	0.0069	85.69	0.0032
Random Forest	88.38	0.0056	90.62	0.0074	<b>94.05</b>	0.0033
Gradient Boosting	81.27	0.0098	87.47	0.0062	91.50	0.0065
DNN	77.16	0.0093	85.47	0.0087	89.65	0.0083

**TABLE 7.** All features accuracy/k fold = 5 on FiSC dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic regression	86.31	0.0042	93.17	0.0078	94.45	0.0116
K-nearest neighbors	86.84	0.0109	93.22	0.0060	93.42	0.0110
(Kernel) SVM	87.51	0.0070	93.94	0.0066	<b>95.23</b>	0.0082
Decision tree	84.26	0.0123	91.89	0.0071	91.06	0.0084
Random Forest	87.41	0.0089	94.48	0.0076	94.56	0.0061
Gradient Boosting	86.90	0.0098	93.95	0.0058	94.39	0.0061
DNN	85.78	0.0081	93.25	0.0086	94.45	0.0142

**TABLE 8.** All features accuracy/k fold = 5 on FiSC-aug dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic regression	79.19	0.0040	88.00	0.0066	92.59	0.0098
K-nearest neighbors	95.96	0.0026	95.89	0.0030	<b>96.91</b>	0.0033
(Kernel) SVM	86.67	0.0042	89.37	0.0056	94.26	0.0039
Decision tree	76.22	0.0045	84.26	0.0076	88.58	0.0061
Random Forest	95.31	0.0058	95.88	0.0026	96.49	0.0016
Gradient Boosting	89.99	0.0074	93.82	0.0039	95.92	0.0014
DNN	85.73	0.0070	88.90	0.0077	93.08	0.0086

Windows 10 64-bit operating system along with a GPU. We used Python and TensorFlow over Keras [44] for the proposed algorithm implementation, which is a complete environment for ML and deep learning algorithms. For evaluating the ML and deep learning algorithms, we set cross-

validation splitting over the dataset into five stratified folds for training and testing the ML classifiers described in the previous section and executed in 30 rounds. For reflection removal, we set empirically the discrimination criteria for the static test  $\delta > 200$ , and for the variable test, we set the constant  $C = 1.15$  and obtain  $\delta$  dynamically per image from (10), this dynamical  $\delta$ , we call “Variable  $\delta$ ” in the experimental results presentation.

### C. RLH FEATURES EVALUATION

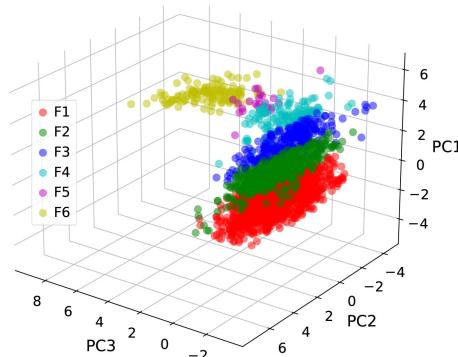
In our first experiments, we evaluated the feature selection. We selected skin-related features from the color spaces studied [22], which uses Chroma, L\*, and Hue channels to explain the color differences over these CS components. We use the Red channel instead of Chroma because Chroma is a component that sub-samples RGB CS. Also, the Red channel is empirically the most representative channel for human skin. The selected features extracted from images are represented by the feature vector  $X = [r, L, H]$ . We plotted the dataset for visualizing the stratum of each class presented in Fig. 6. The accuracy results of the extracted features are described in Table 5 and 6 using FiSC Dataset and FiSC Aug-Dataset, respectively. As it is seen in the Table 5 and 6, the accuracy reaches over 90% in some classifiers. However, with these experiments we cannot prove that [rLH] features are the most influencing features. Moreover the experiments over the full 9D feature vector [r,g,b,L,A,B,H,S,V] show that the accuracy is not near to the accuracy shown using the same classifiers with it. The results for FiSC and FiSC-Aug datasets are described in Table 7 and 8 respectively.

**TABLE 9.** Covariance matrix raw FiSC dataset.

Color space	Channel 1	Channel 2	Channel 3
RGB	<b>0.80781176</b>	<b>0.15127411</b>	<b>0.02731906</b>
LAB	0.06823429	0.00298513	0.00187477
HSV	0.00009767	0.00004999	0.00002111

### D. INFLUENCING FEATURES IDENTIFICATION

To find the most influencing features aiming for a dimensionality reduction, we used the proposed feature engineering approach to identify the potential influencing features without affecting the accuracy. The covariance matrix among the elements in the 9D feature vector is calculated as in (13) over the raw dataset, to check which representative features on the feature vector maximize the variance. The results in the covariance matrix that have the maximum variance are on RGB CS features as shown in Table 9. Having the potential influencing features that are the values that maximize the variance, we normalized the dataset. Over this normalized dataset we calculated the eigenvalues matrix with (15), and the results are as shown in Table 10. To prepare the dataset, we normalized the feature channels (Table 4) in the raw dataset. Therefore, we will use two terms for the FiSC Dataset, i.e., Normalized FiSC Dataset and Raw FiSC Dataset.



**FIGURE 7.** PCA normalized dataset plot showing the stratum for each class.

**TABLE 10.** Eigenvalues normalized FiSC dataset.

Color space	Channel 1	Channel 2	Channel 3
RGB	<b>4.22213066</b>	<b>3.2580152</b>	<b>0.86288798</b>
LAB	0.61433308	0.02687595	0.01687908
HSV	0.00020571	0.00107006	0.00089174

**TABLE 11.** PCA eigenvalues (EV) and variance (V).

PCA	PC1	PC2	PC3
Eigenvalue (EV)	<b>4.22213066</b>	<b>3.2580152</b>	<b>0.86288798</b>
Variance (V)	0.50606659	0.39050725	0.10342616

Then we analyzed the 9D feature vector of the normalized dataset using PCA and having the eigenvalues and covariance matrix in Table 11, we found that the eigenvalues from PCA components are the same from RGB CS. By definition PCA makes projections using linear transformations over the data into a new coordinate system through eigendecomposition of the covariance matrix. It can be inferred through this definition that the eigenvalues that maximize the variance are the ones representing the most influencing features. Since eigenvalues are the same for PCA over the normalized dataset in Table 11 and normalized dataset for RGB CS in Table 9, it means that they have exactly the same minimal and characteristic polynomials. Thus, we have identified the most influencing features and then it is safe to perform a dimensionality reduction. These influencing features are clustered in what we call the dominant areas as illustrated in Fig. 7 for PCA.

#### E. INFLUENCING FEATURES EVALUATION

We evaluated using the feature vector  $X = [r \ g \ b]$  which contains the most influencing features using the same settings of rLH and the results are shown in Table 12 and 13. The results are clearly near the resultant values of the 9D feature vector over the classifiers.

#### F. COMPARISON WITH STATE-OF-THE-ART

With the unified value per channel, we selected the input values from LAB CS for testing ITA, and input these values in the same way as [10] and [12], the results of applying ITA

**TABLE 12.** rgb accuracy. k fold = 5 on FiSC dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic regression	85.79	0.0038	93.07	0.0072	93.81	0.0118
K-nearest neighbors	85.47	0.0064	93.37	0.0078	94.21	0.0078
(Kernel) SVM	86.28	0.0064	93.86	0.0038	<b>94.53</b>	0.0132
Decision tree	83.09	0.0113	91.27	0.0061	90.36	0.0084
Random Forest	84.79	0.0119	94.16	0.0067	94.14	0.0041
Gradient Boosting	85.30	0.0038	94.15	0.0092	94.20	0.0046
DNN	85.26	0.0058	93.06	0.0076	94.19	0.0115

**TABLE 13.** rgb accuracy. k fold = 5 on FiSC-aug dataset.

Method	No Filtered		Static $\delta$		Variable $\delta$	
	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$	% $\mu$	$\pm\sigma^2$
Logistic regression	73.80	0.0032	87.50	0.0064	91.32	0.0086
K-nearest neighbors	91.14	0.0066	94.55	0.0060	<b>97.00</b>	0.0035
(Kernel) SVM	79.89	0.0098	88.18	0.0072	92.21	0.0060
Decision tree	72.77	0.0076	83.87	0.0063	88.63	0.0062
Random Forest	91.27	0.0029	94.72	0.0079	96.71	0.0028
Gradient Boosting	85.09	0.0049	92.13	0.0087	95.31	0.0052
DNN	81.35	0.0074	88.97	0.0080	92.07	0.0103

**TABLE 14.** Comparison with controlled environment dataset. NFA (No Filter Augmented),  $\delta_{VA}$  (Variable  $\delta$  Augmented).

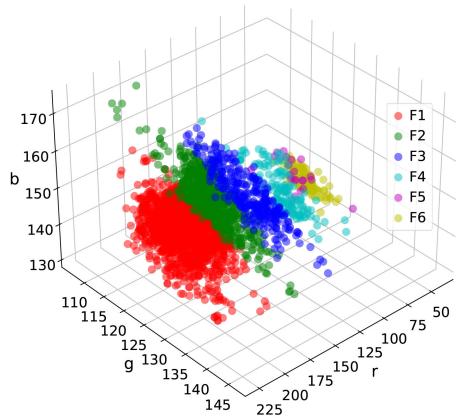
Method	No Filter %	Variable $\delta$ %	NFA %	$\delta_{VA}$ %
ITA (Groh Kinyanjui [10])	10.56	18.12	20.62	40.38
ITA (Groh Empirical [10])	72.32	74.17	33.33	31.36
ITA (Krishnapriya [11], [12])	24.73	23.13	19.54	16.81
ML-FiSC (Ours)	<b>85.47</b>	<b>94.21</b>	<b>91.14</b>	<b>97.00</b>

**TABLE 15.** 30 rounds ML-FiSC KNN mean results varying number of K folds with number of test results (NTR).

K folds	Accuracy%	Precision%	Recall%	F1 Score%	NTR
7	97.10	97.12	97.10	97.10	210
9	97.16	97.18	97.16	97.16	270
11	97.28	97.30	97.28	97.28	330
15	97.12	97.16	97.13	97.13	450
17	97.21	97.25	97.22	97.21	510

to our dataset are presented in Table 14 where we compare our results too. Since our dataset is different than the one used [10] and [12], we understand that results may vary in some degree. Groh Empirical with our dataset without augmentation performed similar accuracy as the reported in their work, however in their work they presented this results with one-tone difference ( $\pm 1$  class) relaxation and we are performing our experiments without relaxation. Along with the differences in datasets, Krishnapriya et. al. [12] accuracy is poor due to the adaption of ITA parameters to their full faces dataset, and they use color-corrected images to classify the Fitzpatrick skin tone. We selected the KNN classifier to do this comparison since it is the best result for augmented data having an accuracy of 97% over the augmented dataset. This high accuracy can be explained due to the nature of the algorithm, when we identified the most influencing features, we also aimed to identify the dominant areas, as seen in Fig. 8 each dominant area is clustered and has the strata well defined, that validates the effectiveness of our filter and reflection removal techniques. When KNN searches for the nearest neighbor in clustered data, the accuracy is high. To verify that the model is not overfitting we performed an

extra experiment varying the number of k folds having the results presented in Table 15.



**FIGURE 8.** [rgb] extracted from dataset, a plot showing the stratum for each class.

## V. IMPLICATIONS AND LIMITATIONS

Our proposed methodology (FiSC) presents several significant outcomes and implications:

### A. NOVEL CLASSIFICATION TECHNIQUE

FiSC introduces an innovative approach to Fitzpatrick scale-based image classification, leveraging state-of-the-art machine learning algorithms. By utilizing specialized filters to detect and remove skin surface imperfections, FiSC achieves enhanced measurement accuracy compared to existing methods.

### B. SKIN FILTER AND REFLECTION REMOVAL

The development of a skin filter capable of accurately identifying skin surfaces in images, even in the presence of darker pixels caused by skin texture, addresses a critical challenge in image-based skin tone classification. Additionally, the novel reflection removal technique effectively mitigates biases introduced by skin sebum reflections, further improving classification accuracy.

### C. FEATURE ENGINEERING AND DIMENSIONALITY REDUCTION

Through comprehensive feature engineering and dimensionality reduction techniques, FiSC identifies the most influential features for Fitzpatrick scale classification. This approach enhances the interpretability of classification results and facilitates more effective utilization of machine learning models.

### D. EXPERIMENTAL VALIDATION

Extensive experiments conducted using the FiSC framework demonstrate significant improvements in classification accuracy, achieving up to 97% accuracy across various skin types. These results underscore the efficacy of the

proposed methodology in accurately classifying skin tones in controlled environments.

The limitation of our study stems from constraints in data acquisition. The FiSC framework relies on image datasets obtained via specialized skin analyzer devices, potentially restricting the availability and diversity of training data. Further efforts are needed to expand the dataset, ensuring comprehensive coverage of diverse skin types and environmental conditions. While FiSC demonstrates commendable accuracy in controlled settings, its performance in real-world clinical scenarios with variable lighting conditions and image quality requires evaluation. Additional validation studies across diverse clinical settings are imperative to assess the robustness and applicability of the proposed methodology. Nonetheless, the methodology can potentially be replicated using general datasets containing high-quality images, provided suitable filters for accurate skin detection are employed.

## VI. CONCLUSION AND FUTURE WORK

In this study, we introduced the FiSC (Fitzpatrick Scale Classification) framework, providing a robust and accurate solution for Fitzpatrick scale-based skin tone classification. Our approach enhances measurement accuracy by applying specialized filters to remove skin surface attributes, such as wrinkles and pores, using a controlled dataset from a professional skin analyzer device. We developed a 9-dimensional feature vector and employed feature engineering to identify the most influential features and dominant areas, enabling dimensionality reduction. This reduction resulted in lighter inputs and decreased complexity for Machine Learning (ML) and Deep Learning (DL) models by minimizing parameters. Extensive classification experiments with various ML algorithms demonstrated that FiSC achieved up to 97% accuracy, surpassing existing state-of-the-art methods. Additionally, our validation showed that the reduced feature set maintained performance comparable to the full nine-dimensional vector, confirming the effectiveness of our dimensionality reduction strategy. These findings have significant implications for applications in cosmetic product development, personalized skincare, and improving the fairness and accuracy of computer vision algorithms across diverse populations. By enhancing skin tone classification accuracy and reducing model complexity, the FiSC framework contributes to more inclusive and effective technological solutions in the healthcare and beauty industries.

In future work, we will focus on addressing the current limitations of our study to further enhance the applicability and effectiveness of the FiSC framework in real-world healthcare settings. Additionally, we plan to apply our methodology to other general datasets to identify influencing features and dominant areas, thereby optimizing dimensional reduction and improving the efficiency and accuracy of Fitzpatrick scale classification.

## ACRONYMS

<b>AHE</b>	Adaptive Histogram Equalisation.
<b>CLAHE</b>	Contrast Limited Adaptive Histogram Equalization.
<b>CS</b>	Color Space.
<b>DP</b>	Dermo Prime.
<b>FiSC</b>	Fitzpatrick Scale based Classification utilizing ML techniques.
<b>HSV</b>	Hue, Saturation, and Value.
<b>ITA</b>	Individual Typology Angle.
<b>LAB</b>	CIELAB.
<b>ML</b>	Machine Learning.
<b>PCA</b>	Principal Component Analysis.
<b>PI</b>	Pigmentation Index.
<b>ROI</b>	Region of Interest.
<b>SPAD</b>	Single-photon Avalanche Diode.
<b>STM</b>	Skin Tone Meter.
<b>UV</b>	Ultraviolet.

## REFERENCES

- [1] O. Arosarena, "Options and challenges for facial rejuvenation in patients with higher Fitzpatrick skin phototypes," *JAMA Facial Plastic Surgery*, vol. 17, no. 5, pp. 358–359, Sep. 2015.
- [2] C. Herzman, S. D. Walter, L. From, and A. Alison, "Observer perception of skin color in a study of malignant melanoma," *Amer. J. Epidemiol.*, vol. 126, no. 5, pp. 901–911, Nov. 1987.
- [3] P. J. Bevan and A. Atapour-Abarghouei, "Detecting melanoma fairly: Skin tone detection and debiasing for skin lesion classification," 2022, *arXiv:2202.02832*.
- [4] P. Aggarwal and F. A. Papay, "Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations," *J. Dermatolog. Treatment*, vol. 33, no. 4, pp. 2257–2262, May 2022.
- [5] M. Groh, O. Badri, R. Daneshjou, A. Koochek, C. Harris, L. R. Soenksen, P. M. Doraiswamy, and R. Picard, "Deep learning-aided decision support for diagnosis of skin disease across skin tones," *Nature Med.*, vol. 30, no. 2, pp. 573–583, Feb. 2024.
- [6] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Arch. Dermatol.*, vol. 124, no. 6, p. 869, Jun. 1988.
- [7] C. Ash, G. Town, P. Bjerring, and S. Webster, "Evaluation of a novel skin tone meter and the correlation between Fitzpatrick skin type and skin color," *Photon. Lasers Med.*, vol. 4, no. 2, pp. 177–186, Jan. 2015.
- [8] J. C. Lester, J. L. Jia, L. Zhang, G. A. Okoye, and E. Linos, "Absence of images of skin of colour in publications of COVID-19 skin manifestations," *Brit. J. Dermatol.*, vol. 183, no. 3, pp. 593–595, Sep. 2020.
- [9] K. Bahmani, R. Plesh, C. Sahu, M. Banavar, and S. Schuckers, "SREDS: A dichromatic separation based measure of skin color," in *Proc. IEEE Int. Workshop Biometrics Forensics (IWB)*, May 2021, pp. 1–6.
- [10] M. Groh, C. Harris, L. Soenksen, F. Lau, R. Han, A. Kim, A. Koochek, and O. Badri, "Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1820–1828.
- [11] K. Krishnapriya, M. C. King, and K. W. Bowyer, "Analysis of manual and automated skin tone assignments for face recognition applications," 2021, *arXiv:2104.14685*.
- [12] K. S. Krishnapriya, G. Pangelinan, M. C. King, and K. W. Bowyer, "Analysis of manual and automated skin tone assignments," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 429–438.
- [13] S. Del Bino, J. Sok, E. Bessac, and F. Bernerd, "Relationship between skin response to ultraviolet exposure and skin color type," *Pigment Cell Res.*, vol. 19, no. 6, pp. 606–614, Dec. 2006.
- [14] C. Schumann, G. O. Olanubi, A. Wright, E. P. Monk, C. Heldreth, and S. Ricco, "Consensus and subjectivity of skin tone annotation for ML fairness," in *Proc. 37th Int. Conf. Neural Inf. Process. Syst.*, New Orleans, LA, USA: Red Hook, NY, USA: Curran Associates Inc., 2023, pp. 30319–30348.
- [15] T. Barrett, Q. Chen, and A. Zhang, "Skin deep: Investigating subjectivity in skin tone annotations for computer vision benchmark datasets," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2023, pp. 1757–1771.
- [16] V. R. Weir, K. Dempsey, J. W. Gichoya, V. Rotemberg, and A.-K.-I. Wong, "A survey of skin tone assessment in prospective research," *npj Digit. Med.*, vol. 7, no. 1, p. 191, Jul. 2024.
- [17] J. B. Dawson, D. J. Barker, D. J. Ellis, J. A. Cotterill, E. Grassam, G. W. Fisher, and J. W. Feather, "A theoretical and experimental study of light absorption and scattering by in vivo skin," *Phys. Med. Biol.*, vol. 25, no. 4, pp. 695–709, Jul. 1980.
- [18] P. H. Andersen and P. Bjerring, "Spectral reflectance of human skin in vivo," *Photodermatol. Photoimmunol. Photomedicine*, vol. 7, no. 1, pp. 5–12, Feb. 1990.
- [19] P. Bjerring and P. H. Andersen, "Skin reflectance spectrophotometry," *Photo-Dermatol.*, vol. 4, no. 3, pp. 167–171, Jun. 1987.
- [20] T. Dwyer, K. Müller, L. Blizzard, R. Ashbolt, and G. W. M. Phillips, "The use of spectrophotometry to estimate melanin density in Caucasians," *Cancer Epidemiol. Biomarkers Prevention: A Publication Amer. Assoc. Cancer Res., Cosponsored Amer. Soc. Preventive Oncol.*, vol. 7, p. 70, Jun. 1997.
- [21] J. A. Viator, J. Komadina, L. O. Svaasand, G. Aguilar, B. Choi, and J. S. Nelson, "A comparative study of photoacoustic and reflectance methods for determination of epidermal melanin content," *J. Investigative Dermatol.*, vol. 122, no. 6, pp. 1432–1439, Jun. 2004.
- [22] I. L. Weatherall and B. D. Coombs, "Skin color measurements in terms of CIELAB color space values," *J. Investigative Dermatol.*, vol. 99, no. 4, pp. 468–473, Oct. 1992.
- [23] Cortex Technology Aps, Cortex, Palo Alto, CA, USA, 2022.
- [24] N. A. bin Abdul Rahman, K. C. Wei, and J. See, "RGB-H-CbCr skin colour model for human face detection," *Fac. Inf. Technol., Multimedia Univ.*, vol. 4, Jan. 2007.
- [25] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognit. Lett.*, vol. 41, pp. 3–13, May 2014.
- [26] R. F. Rahmat, T. Chairunnisa, D. Gunawan, and O. S. Sitompul, "Skin color segmentation using multi-color space threshold," in *Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Aug. 2016, pp. 391–396.
- [27] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [28] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2018.
- [29] T. Song, J. Feng, L. Luo, C. Gao, and H. Li, "Robust texture description using local grouped order pattern and non-local binary pattern," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 189–202, Jan. 2021.
- [30] A. Sarkar, A. L. Abbott, and Z. Doerzaph, "Universal skin detection without color information," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 20–28.
- [31] L. Zhou, Z. Liu, and X. He, "Face parsing via a fully-convolutional continuous CRF neural network," 2017, *arXiv:1708.03736*.
- [32] Dermo-Prime Plus, Chowis, Yongin, South Korea, Jan. 2022.
- [33] C. Connolly and T. Fleiss, "A study of efficiency and accuracy in the transformation from RGB to CIELAB color space," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 1046–1048, Jul. 1997.
- [34] H. Lidong, Z. Wei, W. Jun, and S. Zebin, "Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement," *IET Image Process.*, vol. 9, no. 10, pp. 908–915, Oct. 2015.
- [35] K. J. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. USA: Academic Press Professional, Inc., 1994, pp. 474–485.
- [36] J.-Y. Kim, S.-H. Kim, and G.-J. So, "Evaluation of perceived depth model with change of hue information," *Int. J. Comput. Commun. Eng.*, vol. 3, no. 1, pp. 75–80, 2014.
- [37] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398. Hoboken, NJ, USA: Wiley, 2013.
- [38] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric discrimination: Consistency properties," *Int. Stat. Rev./Revue Internationale de Statistique*, vol. 57, no. 3, p. 238, Dec. 1989.
- [39] J. R. Quinlan, "Simplifying decision trees," *Int. J. Man-Mach. Stud.*, vol. 27, no. 3, pp. 221–234, Sep. 1987.

- [40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [41] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [42] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [43] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [44] N. Ketkar and N. Ketkar, "Introduction to Keras," in *Deep Learning With Python: A Hands-on Introduction*. Berkeley, CA, USA: Apress., 2017, pp. 97–111. [Online]. Available: [https://doi.org/10.1007/978-1-4842-2766-4\\_7](https://doi.org/10.1007/978-1-4842-2766-4_7)



**GUILLERMO CROCKER GARCIA** received the Bachelor of Science (B.Sc.) degree in systems, informatics, and computer engineering from Universidad Francisco Marroquin, Guatemala, in 2006, and the Master of Science (M.Sc.) degree in information systems and databases from Universidad Galileo, Guatemala, in 2007. Currently, he is pursuing the Ph.D. degree with Kyung-Hee University, South Korea. Throughout his academic and professional career, he has gained valuable experience in industry settings. He has contributed his expertise to companies in South Korea, including R-Spirit, in 2017 and Chowis Company Ltd., since 2018. In these companies, he has developed different products using computer vision, machine learning, and AI. His research interests include cutting-edge areas of computer engineering, including computer vision, machine learning, artificial intelligence, and big data.



**MUHAMMAD NUMAN KHAN** received the bachelor's degree in computer science with a specialization in data science and web engineering from the University of Peshawar, Pakistan, where he developed a strong foundation in software development and data analytics. He is currently pursuing the Ph.D. degree in computer science and engineering with Kyung-Hee University, Global Campus, South Korea.

His Ph.D. research focuses on advancing techniques in big data analytics, distributed computing, and machine learning to address real-world challenges in video retrieval, computer vision, and information systems. His research interests include a broad range of topics, including big data analytics, distributed computing, machine and deep learning, and computer vision. He is particularly focused on the application of these technologies in real-world challenges. He has been actively involved in various research projects. His work seeks to bridge the gap between theoretical advancements and practical applications, particularly in optimizing large-scale data processing frameworks and enhancing intelligent systems.



**AFTAB ALAM** received the M.S. degree in computer science from the University of Peshawar, Pakistan, and the Ph.D. degree from the Data and Knowledge Engineering Laboratory, Department of Computer Engineering, Kyung Hee University (KHU), Global Campus, South Korea, under Prof. Young-Koo Lee's supervision in the domain of big data analytics and industrial scale-out systems designing. He is currently a Data Manager with Quaid-i-Azam University, Islamabad, and a Visiting Professor with NUML University, Islamabad. He has been doing research and industrial development in computer science for more than the last 14 years. He got another chance as a Researcher in the domain of digital well-being under the supervision of Prof. Dr. Raian Ali with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. His research work, during his Ph.D. and postdoctoral at KHU, was in the area of shared-nothing architecture-based distributed computing solutions for big data, i.e., designing industrial-scale distributed computing architectures for big data analytics in the cloud and subsequently, addressing the research issues and challenges being identified in the designed big data systems. His research expertise includes experience in ranking experts on online knowledge information-seeking platforms, utilizing distributed computing technologies for big data analysis with advanced machine learning techniques, conducting complex event analysis in multi-stream environments using semantic web technologies, and exploring digital well-being.



**JOSUE OBREGON** received the Ph.D. degree in industrial engineering from Kyung Hee University, in 2020. He is currently an Assistant Professor with the Department of Industrial Engineering, Seoul National University of Science and Technology (SeoulTech), South Korea. His research interests include explainable artificial intelligence, smart manufacturing, and smart energy.



**TAMER ABUHMED** (Senior Member, IEEE) received the Ph.D. degree in information and telecommunication engineering from Inha University, in 2012. He is currently an Associate Professor with the College of Computing and Informatics, Sungkyunkwan University, South Korea. His research interests include information security, trustworthy and adversarial machine learning, and expert systems for medical applications.



**EUI-NAM HUH** received the B.S. degree from Busan National University, South Korea, the master's degree in computer science from The University of Texas at Austin, USA, in 1995, and the Ph.D. degree from Ohio University, USA, in 2002. He is currently a Professor with the Department of Computer Science and Engineering, Kyung-Hee University, South Korea. His research interests include cloud computing, the Internet of Things, future internet, distributed real-time systems, distributed learning, cloud brain, and security. He is a member of the Review Board of the National Research Foundation of Korea. He has also served many community services for ICCSA, WPDRTS/IPDPS, APAN Sensor Network Group, ICUIMC, ICONI, APIC-IST, ICUFN, and SoICT as various types of chairs. He is the Vice-chairperson of the Cloud/Bigdata Special Technical Group of TTA and an Editor of ITU-T SG13 Q19.