Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis

Liang, Yunfeng; Sun, Lei; Ser, Wee; Lin, Feng; Thng, Steven Tien Guan; Chen, Qiping; Lin, Zhiping

2018

Liang, Y., Sun, L., Lin, F., Thng, S. T. G., Chen, Q., & Lin, Z. (2018). Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis. *Computers in Biology and Medicine*, 99, 123-132.
doi:10.1016/j.combiomed.2018.05.026

<https://hdl.handle.net/10356/136676>

<https://doi.org/10.1016/j.combiomed.2018.05.026>

© 2018 Elsevier Ltd. All rights reserved. This paper was published in *Computers in Biology and Medicine* and is made available with permission of Elsevier Ltd.

Downloaded on 26 Apr 2023 17:27:56 SGT

Classification of Non-Tumorous Skin Pigmentation Disorders Using Voting Based Probabilistic Linear Discriminant Analysis

Yunfeng Liang^a, Lei Sun^b, Wee Ser^c, Feng Lin^d, Steven Tien Guan Thng^e,
Qiping Chen^e, Zhiping Lin^{c,*}

^a*Interdisciplinary Graduate School, Nanyang Technological University, Singapore.*

^b*School of Information and Electronics, Beijing Institute of Technology, Beijing, PR China.*

^c*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.*

^d*School of Computer Science and Engineering, Nanyang Technological University, Singapore.*

^e*National Skin Centre, 1 Mandalay Road, Singapore.*

Abstract

Non-tumorous skin pigmentation disorders can have a huge negative emotional impact on patients. The correct diagnosis of these disorders is essential for proper treatments to be instituted. In this paper, we present a computerized method for classifying five non-tumorous skin pigmentation disorders (i.e., freckles, lentigines, Hori's nevus, melasma and nevus of Ota) based on probabilistic linear discriminant analysis (PLDA). To address the large within-class variance problem with pigmentation images, a voting based PLDA (V-PLDA) approach is proposed. The proposed V-PLDA method is tested on a dataset that contains 150 real-world images taken from patients. It is shown that the proposed V-PLDA method obtains significantly higher classification accuracy (4% or more with $p < 0.001$ in the analysis of variance (ANOVA) test) than the original PLDA method, as well as several state-of-the-art image classification methods. To the authors' best knowledge, this is the first study that focuses on the non-tumorous skin pigmentation im-

*Corresponding author. School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798.

Email address: ezplin@ntu.edu.sg (Zhiping Lin)

age classification problem. Therefore, this paper could provide a benchmark for subsequent research on this topic. Additionally, the proposed V-PLDA method demonstrates promising performance in clinical applications related to skin pigmentation disorders.

Keywords: V-PLDA, non-tumorous, skin pigmentation disorders, image classification, large within-class variance

1. Introduction

Over the past few decades, digital imaging techniques have been widely applied to assist the diagnosis of skin pigmentation disorders [1, 2, 3]. However, most studies focus on the detection and classification of skin tumors, such as melanoma [4, 5] and pigmented basal cell carcinoma (BCC) [6, 7]. While such studies are very important because skin tumors are likely to be cancers and threaten the health of the patients [8], some attention has recently been paid to developing computerized methods for the severity assessment of non-tumorous skin pigmentation disorders [9, 10]. Non-tumorous skin conditions, such as melasma, lentigines, Hori’s Nevus, and so on, mainly appear on the face and they are regarded as unsightly, causing much emotional stress to the sufferers [11]. Because non-tumorous skin pigmentation disorders are not fatal, many of these patients seek help from beauty salons and aestheticians instead of dermatologists. However, without the professional knowledge of skin pigmentation treatments [12, 13], non-professionals may make a wrong diagnosis [14, 15] and recommend options such as lasers and chemical peels, which could result in adverse effects on the patients [16]. With this background, the objective of this work is to provide the patients with a computerized tool that can easily and reliably diagnose non-tumorous skin pigmentation disorders based on their own facial image photographs. To the author’s best knowledge, this is the first study dealing with the non-tumorous skin pigmentation classification problem.

It is important to point out that diagnosis rules for non-tumorous skin pigmentation disorders differ greatly from those for their tumorous counterparts due to the appearance differences in their associated skin images. Tumorous conditions like melanoma, and BCC usually present as a distinct single lesion and a detailed image analysis of the lesion is carried out to look for specifics like size, shape, color, and structure. These prominent features are summarized to form some precise rules, for example, the ABCD rule

Figure 1: Example images of melasma patients. The eyes are covered manually to protect privacy. It can be observed that the melasma pigmentation patches on each face make large difference in the appearance.

[17, 18], for skin tumor image classification. In contrast, the features of the non-tumorous skin pigmentation disorders usually affect the whole face and are bilateral. The pigmentation appearance varies a lot in terms of size, color and shape even for the ones in the same class. An example is illustrated in Figure 1 for melasma, a common non-tumorous pigmentation. Due to the above-mentioned differences, methods specifically designed for the classification of tumorous pigmentation disorders may not be suitable for solving the classification problem of non-tumorous pigmentation disorders. Therefore, it is necessary to look for other more suitable image classification methods.

Many image classification methods have been developed in the past decades and they could potentially be used for non-tumorous skin image classification. These methods can be broadly classified into local and global approaches [19, 20]. In the local approach, distinguished local patches in a given image are firstly marked by region detectors. The features are then extracted from these patches using different feature descriptors, such as scale invariant feature transform (SIFT) [21] and local binary patterns (LBP) [22], and so on, to represent the image and do the classification. While local methods could be used for skin image classification, the local region selection and description need to be carefully designed. For the real-world non-tumorous skin pigmentation image set that we have at hand, we observe that the classification accuracy is not high due to the large within-class variance in the image set.

An alternative to the local approach is the global approach, which uses the whole image as an input so that the local region detection process is

not required [23, 24]. To transform the high dimensional input image into a class label, the global methods project the data onto a suitable lower dimensional space. For example, the Fisher’s linear discriminant analysis (LDA) based methods [23, 25] project the input images onto a space with minimized within-class variance. Another well-known linear discriminant analysis model is the LDA incorporated with probability theory to form the probabilistic linear discriminant analysis (PLDA) model [26]. The PLDA is superior to the LDA since it models the categorical variances in images more accurately. However, one problem with some of these global methods, such as LDA and PLDA, is that they are more suitable for problems with small or moderate within-class variance because the image patterns need to be closely aligned [20]. To handle large within-class variance, PLDA is combined with a tied analysis (tied-PLDA) [27, 28] so that the categorical (class) images are further divided into subclasses. This allows the large within-class variance of the whole class to be replaced by several smaller within-class variances with respect to the divided subclasses. It should be pointed out that the tied-PLDA model requires prior information in the training process to divide the data into subclasses. For example, in [27], the tied-PLDA model requires the pose angle of each image in the training procedure. However, in our non-tumorous skin pigmentation image set, no such prior knowledge is available, which may typically be the case with other skin pigmentation image sets.

In this paper, we propose an effective method for non-tumorous skin pigmentation classification, with an emphasis on tackling the large within-class variance problem. Specifically, a voting based PLDA (V-PLDA) method is developed by incorporating a voting scheme into an approximate PLDA method. We first show that for a dataset with large within-class variance, the original PLDA model could be approximated with an individual matching process. This is further enhanced via a voting scheme. The main advantage of the proposed V-PLDA method is that it excludes samples that have small joint probabilities with the testing sample while utilizing those samples that have large joint probabilities with the testing sample, which makes the final classification decision more robust.

The rest of the paper is organized as follows. In Section 2, the PLDA method is briefly reviewed. Section 3 discusses the problem of large within-class variance, the individual matching process, and the proposed V-PLDA method. Simulations are also carried out in this section to verify the effectiveness of the proposed V-PLDA method. Section 4 applies the proposed V-PLDA method to the real-world non-tumorous skin pigmentation image

set and compares the results with both the original PLDA method and several state-of-the-art local and global-based image classification methods. Finally, this paper will conclude in Section 5.

2. PLDA

Probabilistic linear discriminant analysis (PLDA) uses the generative model, which incorporates both within- and between-class variances. In PLDA, the input data are regarded as a linear combination of vectors associated with the within- and between-class covariances [28]:

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad (1)$$

where $\mathbf{x}_{ij}$ is the j th data sample of the i th class, i and j are in the index sets of $[1, 2, \dots, I]$ and $[1, 2, \dots, J]$, respectively. $\boldsymbol{\mu}$ is the mean value of the whole training data. $\mathbf{F}$ and $\mathbf{G}$ denote two matrices whose columns are the basis vectors of the between-class covariance and within-class covariance, respectively. $\mathbf{h}_i$ and $\mathbf{w}_{ij}$ are the corresponding weights of the basis vectors. $\mathbf{h}_i$ represents the between-class character and the row elements remain the same for all the samples from the same class i . In contrast, $\mathbf{w}_{ij}$ represents the within-class character and the row elements change with not only the class label i but also the sample label j . $\boldsymbol{\epsilon}_{ij}$ is a stochastic noise term with a diagonal covariance matrix $\boldsymbol{\Sigma}$. The assembled parameters $\{\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Sigma}\}$ are denoted as $\boldsymbol{\theta}$.

From model (1) and the assumption of Gaussian distributions of the hidden variables $\mathbf{h}_i$ and $\mathbf{w}_{ij}$ [28]:

$$P(\mathbf{h}_i) = \mathcal{N}_{\mathbf{h}_i}(\mathbf{0}, \mathbf{I}), \quad (2)$$

$$P(\mathbf{w}_{ij}) = \mathcal{N}_{\mathbf{w}_{ij}}(\mathbf{0}, \mathbf{I}). \quad (3)$$

We can derive the conditional probability as follows [28]:

$$P(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta}) = \mathcal{N}_{\mathbf{x}_{ij}}(\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \boldsymbol{\Sigma}), \quad (4)$$

where $P(\cdot)$ and $P(\cdot|\cdot)$ denote prior probability and conditional probability, respectively; and, $\mathcal{N}_{\mathbf{a}}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian distribution in $\mathbf{a}$ with mean $\mathbf{m}$ and covariance $\mathbf{C}$. The data distributions in the case of binary class and a two-dimensional data set of this model are illustrated in Figure 2.

Based on (2) - (4), $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$ is the joint probability of different samples that belong to the same class i . $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$

Figure 2: Data distributions of the PLDA model. The two large ellipses represent the covariances of two different classes. The small ellipses inside the large ones are the distributions of individual samples in each class. The between-class center ($\mu + \mathbf{Fh}_i$) and within-class center ($\mathbf{Gw}_{ij}$) are pointed by dashed arrows and solid arrows, respectively. $\mathbf{x}_{ij}$ are the selected representative samples while $\mathbf{x}_t$ denotes the testing sample. The categorical overlapping is highlighted.

is the joint probability of not only the samples of class i but also the testing sample $\mathbf{x}_t$. Consequently, by considering all of the training samples, the probability of the testing sample $\mathbf{x}_t$ belonging to the i th class can be expressed as [28]:

$$P_i = P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t) \prod_{p=1, p \neq i}^I P(\mathbf{x}_{p1}, \mathbf{x}_{p2}, \dots, \mathbf{x}_{pJ}). \quad (5)$$

The classification can be obtained by assigning $\mathbf{x}_t$ to the class i which has the largest probability P_i :

$$C = \arg \max_i P_i, \quad (6)$$

where C is the output class of the testing sample $\mathbf{x}_t$. The parameters in θ can be derived by a Bayesian inference approach, further details can be found in [28, 29].

3. Proposed Method

In the calculation of the joint probability in (5), all of the samples in class i are assumed to contain discriminant information. However, this assumption

is true only when the within-class variance is small or moderate. In other words, all of the samples in the same class would have similar contribution to the joint probability P_i in (5). In contrast, when the within-class variance is large, samples in the same class can have large differences in their contributions to the joint probability and, hence, reduce the performance of the PLDA method. For example, as illustrated in Figure 2, assume that the two selected samples, $\mathbf{x}_{11}$ and $\mathbf{x}_{12}$, and the testing sample $\mathbf{x}_t$ are all from class 1. Apparently, sample $\mathbf{x}_{12}$ would have a larger joint probability with $\mathbf{x}_t$ than that $\mathbf{x}_{11}$ with $\mathbf{x}_t$. Therefore, while the information from $\mathbf{x}_{12}$ suggests that $\mathbf{x}_t$ is likely from class 1, the information from $\mathbf{x}_{11}$ tends to suggest the opposite due to its small joint probability with $\mathbf{x}_t$. By considering both samples $\mathbf{x}_{11}$ and $\mathbf{x}_{12}$, the joint probability $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t)$ in (5) would be small and unreliable for the decision making. Hence, the original PLDA method requires further study in the situation of large within-class variance.

3.1. PLDA as Individual Matching

To understand how the original PLDA model works in the large within-class variance situation, the joint probability $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ} | \mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$ in the PLDA model should be studied first. To derive the detailed expression of the joint probability, the compound form of (1) by considering all the samples in class i together is shown in (7) [28].

$$\begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{iJ} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i1} \\ \mathbf{w}_{i2} \\ \vdots \\ \mathbf{w}_{iJ} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{i1} \\ \boldsymbol{\epsilon}_{i2} \\ \vdots \\ \boldsymbol{\epsilon}_{iJ} \end{bmatrix}, \quad (7)$$

which can be expressed compactly as [28]:

$$\mathbf{x}'_i = \boldsymbol{\mu}' + \mathbf{A}\mathbf{y}_i + \boldsymbol{\epsilon}'. \quad (8)$$

From (8), and by the assumption of Gaussian distribution of the hidden variable $\mathbf{y}_i$ [28]:

$$P(\mathbf{y}_i) = \mathcal{N}_{\mathbf{y}_i}(\mathbf{0}, \mathbf{I}), \quad (9)$$

the conditional probability can be derived as [28]:

$$P(\mathbf{x}'_i | \mathbf{y}_i) = \mathcal{N}_{\mathbf{x}'_i}(\boldsymbol{\mu}' + \mathbf{A}\mathbf{y}_i, \boldsymbol{\Sigma}'), \quad (10)$$

where $\Sigma' = \text{diag}\{\Sigma, \dots, \Sigma\}$.

The marginal distribution is obtained by marginalizing over $\mathbf{y}_i$ [28]:

$$P(\mathbf{x}'_i) = \mathcal{N}_{\mathbf{x}'_i}(\boldsymbol{\mu}', \mathbf{A}\mathbf{A}^T + \Sigma'). \quad (11)$$

Based on (7) and (8), the combined covariance matrix $\mathbf{A}\mathbf{A}^T + \Sigma'$ is expressed as:

$$\mathbf{A}\mathbf{A}^T + \Sigma' = \begin{bmatrix} \mathbf{D} & \mathbf{F}\mathbf{F}^T & \dots & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T & \mathbf{D} & \dots & \mathbf{F}\mathbf{F}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}\mathbf{F}^T & \mathbf{F}\mathbf{F}^T & \dots & \mathbf{D} \end{bmatrix}, \quad (12)$$

where

$$\mathbf{D} = \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T + \Sigma.$$

The combined covariance matrix (12) includes the within-class matrix $\mathbf{G}$ and it provides us with a means to study its impact. When the within-class variance is large, the values of the elements in the within-class matrix $\mathbf{G}$ are large and, hence, the elements of $\mathbf{F}\mathbf{F}^T$ are negligible compared with the corresponding elements in $\mathbf{G}\mathbf{G}^T$. Consequently, the combined matrix $\mathbf{A}\mathbf{A}^T + \Sigma'$ can be approximated by its diagonal elements. This approximation implies that the correlations between different samples in the compound vector $\mathbf{x}'_i = [\mathbf{x}_{i1} \ \mathbf{x}_{i2} \ \dots \ \mathbf{x}_{iJ}]^T$ are small. In this case, the joint probability can be approximated by the product of the individual probabilities:

$$P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}) \approx P(\mathbf{x}_{i1})P(\mathbf{x}_{i2})\dots P(\mathbf{x}_{iJ}). \quad (13)$$

Based on (13), the joint probability of $\mathbf{x}_t$ and the samples in class i can be checked separately, which gives:

$$P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t) \approx \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{q=1, q \neq j}^J P(\mathbf{x}_{iq}). \quad (14)$$

By substituting (13) and (14) into (5), the probability of the testing sample $\mathbf{x}_t$ belonging to the i th class is:

$$\begin{aligned} P_i &\approx \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{q=1, q \neq j}^J P(\mathbf{x}_{iq}) \prod_{p=1, p \neq i}^I P(\mathbf{x}_{p1})P(\mathbf{x}_{p2})\dots P(\mathbf{x}_{pJ}) \\ &= \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{(p,q) \neq (i,j)} P(\mathbf{x}_{pq}) \\ &= \max_j P_{ij}, \end{aligned} \quad (15)$$

where $P_{ij} = P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{(p,q) \neq (i,j)} P(\mathbf{x}_{pq})$ is the probability that $\mathbf{x}_t$ is from the same distribution with only the training sample $\mathbf{x}_{ij}$. Therefore, the derived class label C of the testing sample in (6) is now obtained approximately as:

$$C = \arg \max_i P_i \approx \arg \max_i \left[\max_j P_{ij} \right]. \quad (16)$$

We name the above class assignment process ‘individual matching’ because the training samples are checked individually to obtain the joint probabilities with the testing sample $\mathbf{x}_t$. The approximation process is further evidenced by the following simulation.

An example with synthetic data is used to show the impact of the within-class variance on the classification accuracy when both the individual matching and PLDA methods are applied. Assume that there exist two classes with fixed class centers of $\boldsymbol{\mu}_1 = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_1 = [0.8 \ 0.6]^T$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_2 = [0.19 \ 0.98]^T$. Under this simple set-up, both matrices $\mathbf{F}$ and $\mathbf{G}$ are reduced to a vector as there is only one basis vector in them. To simulate different levels of the within-class variance, $\mathbf{G}$ is set to have different scales by multiplying the direction vector with a scalar ν :

$$\mathbf{G} = \nu \cdot \frac{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}. \quad (17)$$

When the value of ν is larger, the within-class variance will be larger. The basis vector in $\mathbf{G}$ is set to have the direction of $\frac{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}$ so that categorical distributions overlap with each other when the within-class variance is large. The parameter $\mathbf{w}_{ij}$ is randomly generated from a Gaussian distribution defined in (3). The covariance of the noise term is set as $\boldsymbol{\Sigma} = \text{diag}\{0.1, 0.1\}$.

In the simulation, 20 samples from each class are generated according to the distributions that have previously been defined. The 10-fold cross-validation procedure [30] is used for the performance evaluation. The data are randomly divided into 10 equal sized subsets, where nine are used for training and the remaining subset is used for testing. The same procedure is then repeated for the next subset until all 10 subsets have been tested, one by one. The average accuracy of the above 10 tests is obtained. To reduce the randomness, the above training and testing procedure is repeated 10 times randomly. The final average classification accuracy of these 10 runs is used for the performance evaluation.

The classification accuracy results are plotted versus the values of ν in Figure 3, where the accuracy curves can be roughly divided into three parts.

Figure 3: Average classification accuracies of the PLDA method and the individual matching method at different levels of the within-class variance. The two vertical lines separate the plot into three parts.

The first part is the region that the within-class variance is extremely small ($\nu < 0.2$) and, hence, the two classes are well separated. As a result, although the approximation error in (15) is large, the individual matching method can still achieve similar high level of accuracy to that of PLDA. In the second part, when the within-class variance ($0.2 \leq \nu < 1.5$) increases but is not very large, the accuracy difference between PLDA and the individual matching method is large. The increase of the accuracy difference is due to the error that is caused by approximating PLDA with the individual matching method. In the third part, the classification accuracies of the two methods converge to about the same level when the within-class variance is very large ($\nu \geq 1.5$). The convergence of the classification accuracies of the two methods in the third part agrees with the previous observation that PLDA can be approximated by the individual matching when the within-class variance is large enough. The statistical analyses of the results by the paired-sample t-test [31] support this argument. The t-test results show that there is a significant difference in the accuracies obtained from the PLDA and the individual matching process in the second part; for example, at $\nu = 0.5$, $t(999) = 6.7827$, $p < 0.001$. In contrast, there is no significant difference in the third part; for example, at $\nu = 2.0$, $t(999) = 0.2442$, $p > 0.05$.

3.2. Voting Based PLDA

One problem with using the individual matching method to approximate PLDA is that the sample with the largest joint probability P_{ij} in (15) may not always be the desired one. With the large distribution overlapping between classes, it is possible that the sample with the largest joint probability P_{ij} is from a different class than the testing sample. Assume that in Figure 2, $\mathbf{x}_{21}$ is the sample that has the largest joint probability with the testing sample $\mathbf{x}_t$ among all of the samples shown but $\mathbf{x}_{21}$ is in a different class (class 2) from $\mathbf{x}_t$. The individual matching would then obtain a wrong classification decision if it uses $\mathbf{x}_{21}$ alone. We propose to use a voting scheme to deal with this problem and to make the decision process more robust. Instead of considering just a single sample $\mathbf{x}_{ij}$ that has the largest joint probability with $\mathbf{x}_t$, a bag of n samples with the largest joint probabilities with $\mathbf{x}_t$ are used in a voting procedure. The sample selection scheme can be expressed as follows:

$$\mathbb{S} = \{\mathbb{A}(1:n) | \mathbb{A} = \text{sort}(\{P_{ij} : i \in [1, 2, \dots, I]; \\ j \in [1, 2, \dots, J]; \})\}, \quad (18)$$

where $\text{sort}(\cdot)$ is a function that sorts the elements of the input data set in a descending order based on their values. Consequently, the set $\mathbb{S}$ contains the n largest P_{ij} values. In the classification process, each of the n elements is associated with a class i . The final classification decision is made based on a voting procedure using these n samples:

$$C = \arg \max_i |\mathbb{S} \cap \mathbb{P}_i|; \quad \mathbb{P}_i = \{P_{i1}, P_{i2}, \dots, P_{iJ}\}, \quad (19)$$

where $|\cdot|$ is the set cardinality. If more than one class has the same largest number of votes, then the class that contains the largest joint probability P_{ij} is selected. We term this proposed method the voting based PLDA, or V-PLDA in short, to reflect the incorporation of the individual matching approximated PLDA with a voting process.

The voting process of the proposed V-PLDA is illustrated in Figure 4. The selected samples in (18) are those with largest joint probabilities with the testing sample $\mathbf{x}_t$. In this selection, the samples which are less correlated with $\mathbf{x}_t$ and may lead to wrong classification results are not considered in the decision making. Another advantage of the voting process is that it makes the decision more robust. The testing sample $\mathbf{x}_t$ would be misclassified using

Figure 4: The voting sample selection process. Only the ones having high joint probabilities with the testing sample $\mathbf{x}_t$ are selected (circled ones in the figure).

the individual matching when the sample with the largest joint probability is from a different class to the testing sample. This misclassification could happen due to the categorical overlapping caused by the large within-class variance. With the voting process, the result would no longer be determined by just one sample but by several selected samples together. As shown in Figure 4, the number of selected samples in class 1 (a total of 4) is larger than that in class 2 (a total of 3). The classification decision is correct via the voting process regardless of which sample has the largest joint probability. The effectiveness of the proposed V-PLDA method is analyzed using another example with synthetic data, which is described in the following.

When the within-class variance is large, the samples in each class may be divided into subclasses where the within-class variance of each subclass is not large. Although we may not have prior knowledge of making this division in some practical applications, it is feasible to do so for synthetic data because it will enable us to understand the impact of the number of samples in each subclass on the performance of the proposed V-PLDA method. Note that within each subclass, the joint probabilities between samples are large because they usually share some common properties. In general, when a sample in a subclass has a large joint probability with the testing sample $\mathbf{x}_t$, then the other samples in the same subclass would also have a large joint probability with $\mathbf{x}_t$. Hence, when there are more samples in each subclass,

there will be more samples in the voting process that have similar joint probabilities with the testing sample. This makes the final decision more robust. This observation is illustrated in the following simulation.

Table 1: Set-ups for the Subclasses

No. of samples in each subclass m	2	5	10
No. of subclasses in each class K	10	4	2

Figure 5: Average classification accuracies of PLDA and V-PLDA at different subclass sizes.

In the simulation, we use the same set-ups as those in the previous simulation for the individual matching, except that the samples in each class are divided into several subclasses. All of the samples in each subclass share the same within-class character. With the number of samples in each class fixed to 20, the number of samples in each subclass (m) and the number of subclasses (K) can vary in different ways. The three set-ups that are listed in Table 1 are tested. In the simulation, the parameter n is selected from a set of values and the one that achieves the best performance is used. The simulation results are plotted in Figure 5. It can be observed that the classification accuracy increases with the number of samples in each subclass (m). Comparing PLDA with V-PLDA, the simulation clearly shows that

with a fixed number of subclass samples, V-PLDA outperforms PLDA, particularly when the within-class variance is large. To examine the significance of the difference between the accuracies obtained from PLDA and V-PLDA, paired-sample t-tests have been conducted at $\nu = 2.0$ and the results are shown in Table 2. It can be observed that the differences in the accuracies for PLDA and V-PLDA at different m values are significant. From this simulation, we can find that V-PLDA is effective for classification problems with large within-class variance, particularly when the number of samples in each subclass is relatively large. It should be emphasized that V-PLDA differs from tied-PLDA in that the V-PLDA does not require prior knowledge of this subclass division.

Table 2: Results of Paired-Samples t-test Comparing PLDA and V-PLDA with Different m Values

m	PLDA		V-PLDA		t	p
	AC %	SD	AC %	SD		
10	65.27	0.2616	71.93	0.2440	7.482	$1.60e - 13^*$
5	59.75	0.2666	65.48	0.2346	8.748	$9.08e - 18^*$
2	54.40	0.2515	55.95	0.2346	2.408	0.016^*

AC: Average accuracy. SD: Standard Deviation.

*: $N = 1,000$.

In the next section, the V-PLDA method will be applied to a real-world non-tumorous skin pigmentation image classification problem.

4. Classification of Non-tumorous Skin Pigmentation Disorders

4.1. Experimental Set-ups and Image Set

This study focuses on the classification of five types of non-tumorous skin pigmentation disorders based on their appearance, which are: freckles, lentiginos, Hori's nevus, melasma, and nevus of Ota. These five types are selected because they are the most common types [32] and they can be considered as representatives of appearance diagnosis. The other types of non-tumorous skin pigmentation disorders of the face, such as Lichen planus pigmentosus, acanthosis nigricans, maturation pigmentation, are rare and they usually need more professional investigation, such as biopsy, rather than just using appearance information. Therefore, they are not included in this study.

The proposed algorithm is tested on a real-world dataset that consists 30 skin pigmentation images per type. To reduce the impact of external noises (e.g., illumination variations) on the diagnostic accuracy, the data collection process is done under standardized set-ups. A NIKON CMOS sensor digital single-lens reflex camera with 24.2 million pixels is utilized and the output image is saved in the JPEG format. Cosmetic makeup is removed for all of the patients. Flash is used while taking the images to ensure consistent lighting. The distance between the camera and the patient is normally less than 3 meters to record the whole face region, such as the examples shown in Figure 1. A patch that contains the whole or part of the pigmentation region is selected from the recorded image and is pre-processed to have the same dimensionality of $100 \times 100 \times 3$ for the subsequent classification. A representative input image of each type is shown in Figure 6.

In the experiment, the ground truth of classification is provided by dermatologists. Following the same experimental procedure on synthetic data as described in Section 3.1, the repeated 10-fold cross-validation is used to evaluate each algorithm. The voting parameter n used in the experiment is selected as 7 after comparing the cross-validation results of different settings (the details can be found in Section 4.2).

Figure 6: Representative images of the five types of non-tumorous skin pigmentation disorders in the classification problem.

Figure 7: The process of clinical diagnosis of the five types of non-tumorous skin pigmentation disorders.

4.2. Input Feature Selection

With the input image prepared, proper features need to be extracted from them to feed into the V-PLDA model for the classification. The selection of the input feature is based on the clinical golden rules of diagnoses, which are summarized in the flowchart shown in Figure 7. The features mentioned in Figure 7 can be divided into two types: color features and non-color features. The detailed feature selection process of the two types is described in this subsection.

4.2.1. Color feature selection

As illustrated in Figure 7, color is a critical feature that differentiates the selected skin pigmentation disorders. To have a thorough investigation of the color feature, various color spaces, including RGB, rg (normalized RGB), transformed, opponent, Lab, and HSV are tested. These color spaces would reveal the color information of a pixel from many different perspectives [33]. To compare their performance, these color feature vectors are tested using the V-PLDA method with different voting values. The resulting average classification accuracies of the cross-validation test are plotted in Figure 8.

From Figure 8, it can be observed that the choice of color space has a great impact on the performance. In comparison, the HSV and Lab color

Figure 8: Average classification accuracy of V-PLDA model using different color spaces.

spaces achieve better performance than the others. This may happen because the two spaces separate the intensity from the color information and, hence, match the human perception better [34]. The opponent color space performs somewhat better than the RGB color space because it decorrelates the red, green, and blue information in the color, making the channels contain less redundant information. The rg and the transformed color spaces perform poorly in the experiment. They share the similarity of normalizing the input RGB space. This normalization process loses important scale information, so that the classification accuracy is reduced.

Based on the results in Figure 8, we select the HSV color space to extract the color information of the image in the rest of the experiment. The detailed color feature extraction process is shown in Algorithm 1. The output color feature vector $\mathbf{x}_{HSV}$ is then fed into the V-PLDA model that was introduced in Section 3 for the classification. It can also be observed from Figure 8 that the best performance of the algorithm is obtained when the voting number n is around 7. Hence, we fix the voting number as 7 in the rest of the experiment.

4.2.2. Non-color feature selection

Besides the color information, the two non-color features mentioned in Figure 7, namely the speckled/diffused property and the large/small property of the pigmentation region, can be extracted using a Gabor filter bank [35, 36]. In this subsection, the detailed Gabor feature extraction process of an

Algorithm 1 Color feature extraction process.

Input: The input color image in RGB color space, $\mathbf{X}_{RGB} \in \mathbb{R}^{p \times q \times 3}$;

Output: The color feature vector, $\mathbf{x}_{HSV}$;

```
1: for Each pixel  $(x, y)$  in the image do
2:    $[r \ g \ b] \leftarrow \mathbf{X}_{RGB}(x, y, :)$ ;
3:    $c \leftarrow \min(r, g, b)$ ;  $v \leftarrow \max(r, g, b)$ ;
4:    $s \leftarrow \begin{cases} 0 & \text{if } v = 0 \\ (v - c)/v & \text{else} \end{cases}$ 
5:    $h \leftarrow \begin{cases} 60(g - b)/(v - c) & \text{if } v = r \\ 120 + 60(b - r)/(v - c) & \text{if } v = g \\ 240 + 60(r - g)/(v - c) & \text{if } v = b \end{cases}$ 
6:   if  $h < 0$  then
7:      $h \leftarrow h + 360$ 
8:   end if
9:    $h \leftarrow h/360$ 
10:   $\mathbf{X}_{HSV}(x, y, :) = [h \ s \ v]$ ;
11: end for
12: Reshape  $\mathbf{X}_{HSV}$  to the vector form  $\mathbf{x}_{HSV}$ ;
13: return  $\mathbf{x}_{HSV}$ ;
```

input image is explained.

In the spatial domain, a two-dimensional Gabor filter contains two parts, a Gaussian kernel function and a complex sinusoidal plane wave:

$$\Psi_{f,\phi}(x, y) = \frac{f^2}{\pi\gamma\eta} \exp^{-f^2[(x'/\gamma)^2 + (y'/\eta)^2]} \exp^{2\sqrt{-1}\pi f x'}, \quad (20)$$

with:

$$\begin{aligned} x' &= x \cos \phi + y \sin \phi, \\ y' &= -x \sin \phi + y \cos \phi, \end{aligned}$$

where f is the central frequency of the filter, γ and η are the bandwidths along the major and minor axis of the Gaussian function, respectively. ϕ is the angle between the direction of the sinusoidal wave and the x-axis of the spatial domain. A Gabor filter bank contains multiple Gabor filters with different frequencies (f) and orientations (ϕ) to detect different information

Algorithm 2 Non-color feature extraction using Gabor filter bank.

Input: The input image in RGB color space, $\mathbf{X}_{RGB} \in \mathbb{R}^{p \times q \times 3}$;
The Gabor filter bank $\Psi_{f,\phi}$ with user defined U, V values;
Output: The Gabor feature vector of the image, $\mathbf{x}_{Gabor}$;
1: Transform $\mathbf{X}_{RGB}$ to HSV space $\mathbf{X}_{HSV}$ using Algorithm 1;
2: **for** Each frequency $f_u : u \in [0, 1, \dots, U - 1]$ **do**
3: **for** Each orientation $\phi_v : v \in [0, 1, \dots, V - 1]$ **do**
4: Obtain the convolution results $\mathbf{O}_{f_u, \phi_v}$ using (21);
5: Calculate the magnitude of the convolution results $\mathbf{M}_{f_u, \phi_v}$ using (22);
6: Store $\mathbf{M}_{f_u, \phi_v}$ in the collective output matrix $\mathbf{M}_{f, \phi}$.
7: **end for**
8: **end for**
9: Reshape the collative output $\mathbf{M}_{f, \phi}$ to the vector form $\mathbf{x}_{Gabor}$;
10: **return** $\mathbf{x}_{Gabor}$;

in the image. The frequency values are determined by two parameters: the maximum frequency (f_M) and the ratio between two adjacent frequencies (f_r). Therefore, the u th frequency, where $u \in [0, 1, \dots, U - 1]$, is defined as $f_u = f_M / f_r^u$. The orientations are equally separated in the range of $[0 \pi]$, where the v th orientation with $v \in [0, 1, \dots, V - 1]$ can be represented as $\phi_v = v\pi/V$. More details about Gabor filter bank can be found in [35, 37].

Given a Gabor filter bank, the Gabor feature of an input image matrix $\mathbf{X} \in \mathbb{R}^{p \times q \times 3}$ is obtained by first convolving each color channel of $\mathbf{X}$ with different Gabor filters Ψ using symmetric padding. For a Gabor filter with a frequency of f_u and an orientation of ϕ_v , the process can be described as:

$$\mathbf{O}_{f_u, \phi_v}(x, y, z) = \sum_{c=1}^w \sum_{r=1}^w \mathbf{X}(x - c, y - r, z) \Psi_{f_u, \phi_v}(c, r), \quad (21)$$

where $\mathbf{O} \in \mathbb{C}^{p \times q \times 3}$. $\Psi \in \mathbb{C}^{w \times w}$ is the Gabor filter with the size of $w \times w$. Subsequently, the magnitude matrix $\mathbf{M}$ of the convolution output $\mathbf{O}$ is calculated:

$$\mathbf{M}_{f_u, \phi_v}(x, y, z) = \sqrt{\text{Re}(\mathbf{O}_{f_u, \phi_v}(x, y, z))^2 + \text{Im}(\mathbf{O}_{f_u, \phi_v}(x, y, z))^2}, \quad (22)$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ correspond to real and imaginary parts of the input, respectively. Finally, the Gabor feature is obtained by concatenating all of

Table 3: Parameter Selection for the Gabor Filter Bank

Parameter	Symbol	Levels	Values
Eta	η	3	0.5, 1.0, 1.5
Gamma	γ	3	0.5, 1.0, 1.5
Max frequency	f_M	2	1/4, $\sqrt{2}/4$
Frequency ratio	f_r	2	$\sqrt{2}$, 2
Window size	w	4	15, 25, 35, 45

the magnitude matrices $\mathbf{M}_{f_u, \phi_v}$ into the vector form $\mathbf{x}_{Gabor}$, so that it can be fed into the V-PLDA model for classification. The detailed Gabor feature extraction process is summarized in Algorithm 2.

Gabor filters with different wavelet structures are obtained by changing the kernel size of the Gaussian function and the frequency of the sinusoidal wave, which can both be controlled by parameter f in (20). In the experiment, we employed a Gabor filter bank that contains 40 Gabor filters in five frequency levels ($U = 5$) and eight orientations ($V = 8$). Following the method proposed in [37], we fine-tuned five important parameters of the Gabor filter bank using the values listed in Table 3. The selected set-up is $[\eta \ \gamma \ f_M \ f_r \ w] = [1.0 \ 1.5 \ \sqrt{2}/4 \ 2 \ 45]$.

4.2.3. Combination of the input features

With both types of image feature—namely $\mathbf{x}_{HSV}$ and $\mathbf{x}_{Gabor}$ —extracted, it is necessary to combine the two types of information effectively for the subsequent classification process. A common method of feature combination is concatenating them together to form a higher dimensional input vector. However, this method does not work well when the features are not in the same scale. To circumvent the scale difference problem, we propose an alternative to take advantage of the probability framework of the V-PLDA model. Specifically, the information from different features is combined by multiplying the corresponding output probabilities of the V-PLDA model together:

$$P_{ij} = P_{ij}^{HSV} \times P_{ij}^{Gabor}. \quad (23)$$

where P_{ij}^{HSV} is the probability derived from (15) using the HSV feature $\mathbf{x}_{HSV}$ as the input and P_{ij}^{Gabor} is the probability derived with the Gabor feature $\mathbf{x}_{Gabor}$ as the input. The combined probability P_{ij} in (23) will then go through the voting process in (18) and (19) to proceed with the image classification.

Table 4: Accuracy Results and Paired-Sample t-tests Comparing PLDA and V-PLDA with Different Input Features on Image Set

Feature	PLDA		V-PLDA		t	p
	AC %	SD	AC %	SD		
HSV	64.40	0.1254	69.27	0.1020	2.854	0.005*
Gabor	69.93	0.1092	73.67	0.1139	2.433	0.017*
HSV-Gabor	70.20	0.1244	77.33	0.0982	4.082	0.00009*

*: $N = 100$.

4.3. Experimental Results and Discussions

With the combined features, we are now ready to evaluate the proposed V-PLDA method on the skin pigmentation image set introduced in Section 4.1. We start by comparing it with the original PLDA method. The average classification accuracies by PLDA and V-PLDA are listed in Table 4, together with the paired-sample t-test [31] results. The average classification accuracies obtained by the proposed V-PLDA method are higher than those obtained by the original PLDA method with the same input features, while the accuracy differences are significant at a level of $p < 0.05$. Table 4 also shows that with the selected combined features HSV-Gabor, V-PLDA performs better than with HSV or Gabor alone.

One classification example of using the real image set is shown in Figure 9 to illustrate the advantage of the voting process. The seven selected voting samples in (b)–(h) have large joint probabilities with the testing sample in (a). Three samples in (b)–(d) are from a different class to the testing sample. It is important to note that the sample in (b) has the largest joint probability P_{ij} with the testing sample in (a). Hence, the testing sample would have been assigned to the wrong class if we had used the individual matching to approximate the PLDA model. In contrast, the voting process with the seven samples selected leads to the correct classification decision.

To further illustrate the superior performance of the V-PLDA method in dealing with the large within-class variance skin pigmentation problem, the proposed V-PLDA model is compared with several state-of-the-art image classification methods. The methods compared include one from the global approach category (sparse and dense representation (SDR) [24]) and two from the local approach category (locality-constrained linear coding (LLC) [38], intersection kernel support vector machine (IKSVM) [39]). SDR is sim-

Figure 9: (a) The testing sample which is from the class of freckles. (b)–(h) The selected samples in the voting process, where (b)–(d) are the samples from the class of Hori’s nevus and (e)–(h) are from the class of freckles.

ilar to PLDA because it also considers both within- and between-class variances to accurately model the given data. The difference is that SDR is a distance-based method rather than a probabilistic method like PLDA. LLC and IKSVM are representative methods that focus on different aspects of the local classification approaches. While LLC focuses on the feature extraction process by proposing an improved feature coding method, IKSVM focuses on the feature classification process by using a non-linear classifier. In comparison, all the three methods have the ability to adapt to the within-class variance in the data. To make sure that the power of the above methods compared is fully utilized in the comparison, the parameters in those methods are fine tuned in the experiment. In LLC and IKSVM, the dense SIFT descriptor [40] is used after comparing different local descriptors. The experimental results of the proposed V-PLDA method and the methods compared are shown in Table 5.

From Table 5, it can be observed that the proposed V-PLDA model achieves the best performance among all four methods in comparison. The average classification accuracy on our skin pigmentation image set by the proposed V-PLDA is higher than other methods in comparison by about 4% or more. An analysis of variance (ANOVA) [31] has shown that the ef-

Table 5: Results of the Different Methods

Method	AC %	SD
SDR [24]	67.63	0.1141
LLC [38]	71.47	0.1135
IKSVM [39]	73.30	0.1107
V-PLDA (HSV-Gabor)	77.33	0.0982

fect of the classification method used on the resulting accuracy is significant, $F(3, 796) = 27.19$, $p < 0.001$. A Tukey post hoc test [31] has revealed that the average classification accuracy is statistically significantly higher by using the proposed V-PLDA method when compared to the other three methods. These results and analysis show that for the skin pigmentation image set that we have at hand, the proposed V-PLDA model possesses superior discriminating power to the methods that were compared.

While the average classification accuracy gives an overview of the performance, a more detailed classification result of V-PLDA in Table 5 is shown in the confusion matrix in Table 6. It can be found that freckles, melasma, and lentigines can be classified more accurately than Hori’s nevus and nevus of Ota. For Hori’s nevus, over 20% of images are misclassified as freckles. As indicated in Figure 7, the two skin diseases are characterized by small speckles and the main difference is in the color. In the real-world situation, due to the large within-class variance, the slate-brown color of Hori’s nevus differs only slightly from the brown color in the mild symptom case. Hence, images containing mild symptoms of Hori’s nevus might be misclassified as freckles. The same problem exists for nevus of Ota due to the large within-class variance that is caused by the large variation in terms of both color and pigmentation patch size. Consequently, in the case of mild symptoms, the light pigmentation color and small pigmentation patch size of the nevus of Ota image differ only slightly from that of other types of pigmentation images. These images are more likely to be misclassified.

Although the proposed V-PLDA method has obtained better classification accuracy than the original PLDA method and several state-of-the-art image classification methods for the set of non-tumorous skin pigmentation images at hand, there is still room for improvement. First, the available image set is relatively small and this may affect the performance of the classification methods. Therefore, we plan to acquire more images in the future.

Table 6: Confusion Matrix for the V-PLDA Model (%) with Combined HSV and Gabor Features as Its Input

		Predicted Class				
		F	L	H	M	O
True Class	F	84.7	0.7	8.0	6.7	0.0
	L	7.0	80.7	8.7	3.7	0.0
	H	22.3	5.0	67.7	4.3	0.7
	M	6.0	5.0	0.3	82.0	6.7
	O	9.0	6.7	13.0	3.0	68.3

F: freckles; L: lentigines; H: Hori’s nevus;
M: melasma; O: nevus of Ota.

Second, the current image set was collected under a controlled environment, as presented in Section 4.1. To better assist patients and easily and reliably make a diagnosis of non-tumorous skin pigmentation disorders, the algorithm should be further improved to deal with images which are taken under a non-controlled environment. Finally, other features not related to images, such as patients’ ages, could be incorporated to improve the classification accuracy of the proposed V-PLDA method.

5. Conclusion

In this paper, a voting based probabilistic linear discriminant analysis (V-PLDA) method has been proposed to tackle the large within-class variance problem in classifying non-tumorous skin pigmentation images. In contrast to the original probabilistic linear discriminant analysis (PLDA) method, which is more suitable for small or moderate within-class variance, the proposed V-PLDA method alleviates the large within-class variance problem by using only those training samples that have high joint probabilities with the testing sample in the classification process. The selected training samples are utilized by a voting process to obtain the final classification decision. In contrast to the tied-PLDA method, the V-PLDA method does not require any prior information of the training samples.

It has been shown that the proposed V-PLDA method outperforms the original PLDA at different within-class variance levels in experiments using

synthetic data. The V-PLDA method has also been applied to classify a real-world image set containing five types of non-tumorous skin pigmentation disorders: freckles, lentigines, Hori’s nevus, melasma and nevus of Ota. The experimental results have shown that the proposed V-PLDA method achieves significantly better performance than the original PLDA and several state-of-the-art image classification methods when dealing with the given skin image set, which has large within-class variance. We hope that this work will promote more research in helping these patients to diagnose their non-tumorous skin pigmentation problems and to avoid potential improper treatments by non-professionals.

- [1] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, “Techniques and algorithms for computer aided diagnosis of pigmented skin lesions-A review,” *Biomedical Signal Processing and Control*, vol. 39, no. Supplement C, pp. 237 – 262, 2018.
- [2] K. Korotkov and R. Garcia, “Computerized analysis of pigmented skin lesions: A review,” *Artificial Intelligence in Medicine*, vol. 56, no. 2, pp. 69–90, 2012.
- [3] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, “Computational methods for the image segmentation of pigmented skin lesions: A review,” *Computer methods and programs in biomedicine*, vol. 131, pp. 127–141, 2016.
- [4] R. B. Oliveira, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, “Computational methods for pigmented skin lesion classification in images: Review and future trends,” *Neural Computing and Applications*, vol. 27, pp. 1–24, 2016.
- [5] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, “Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold,” *Computers in Biology and Medicine*, vol. 87, pp. 302–310, 2017.
- [6] B. Sciolla, P. Delachartre, L. Cowell, T. Dambry, and B. Guibert, “Improved boundary segmentation of skin lesions in high-frequency 3D ultrasound,” *Computers in Biology and Medicine*, vol. 87, pp. 302–310, 2017.

- [7] S. Y. Chuah, S. I. Tee, W. P. Tan, S. S. J. Lee, S. K. Ng, S. H. Chua, and T. G. S. Thng, "Reflectance confocal microscopy is a useful non-invasive tool in the *in vivo* diagnosis of pigmented basal cell carcinoma in Asians," *Australasian Journal of Dermatology*, vol. 58, no. 2, pp. 130–134, 2017.
- [8] T. L. Diepgen and V. Mahler, "The epidemiology of skin cancer," *British Journal of Dermatology*, vol. 146, no. s61, pp. 1–6, 2002.
- [9] E. Y. Tay, E. Y. Gan, V. W. D. Tan, Z. Lin, Y. Liang, F. Lin, W. Ser, and T. G. Thng, "Pilot study of an automated method to determine melasma area and severity index," *British Journal of Dermatology*, vol. 172, no. 6, pp. 1535–1540, 2015.
- [10] Y. Liang, L. Sun, W. Ser, F. Lin, E. Y. Tay, E. Y. Gan, T. G. Thng, and Z. Lin, "Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment," *Multidimensional Systems and Signal Processing*, vol. 28, pp. 977–994, Jul 2017.
- [11] M. K. A. Basra, R. Fenech, R. M. Gatt, M. S. Salek, and A. Y. Finlay, "The dermatology life quality index 1994–2007: A comprehensive review of validation data and clinical results," *British Journal of Dermatology*, vol. 159, no. 5, pp. 997–1035, 2008.
- [12] N. A. Vashi and R. V. Kundu, "Facial hyperpigmentation: Causes and treatment," *British Journal of Dermatology*, vol. 169, no. s3, pp. 41–56, 2013.
- [13] J. P. Ortonne and D. L. Bissett, "Latest insights into skin hyperpigmentation," in *Journal of Investigative Dermatology Symposium Proceedings*, vol. 13, pp. 10–14, Elsevier, 2008.
- [14] R. F. Wagner, D. Wagner, J. M. Tomich, K. D. Wagner, and D. J. Grande, "Residents' corner: Diagnoses of skin disease: Dermatologists vs. nondermatologists," *The Journal of Dermatologic Surgery and Oncology*, vol. 11, no. 5, pp. 476–479, 1985.
- [15] L. Brochez, E. Verhaeghe, L. Bleyen, and J. M. Naeyaert, "Diagnostic ability of general practitioners and dermatologists in discriminating pig-

- mented skin lesions,” *Journal of the American Academy of Dermatology*, vol. 44, no. 6, pp. 979–986, 2001.
- [16] T. W. Lin, “5 myths about pigmentation debunked by a dermatologist,” 2017. [Online]. Available: <https://thread.zalora.sg/5-myths-about-pigmentation-as-reviewed-by-a-singapore-dermatologist-e5a68fdcccef>.
 - [17] W. Stolz, A. Riemann, A. B. Cognetta, L. Pillet, W. Abmayr, D. Holzel, P. Bilek, F. Nachbar, and M. Landthaler, “ABCD rule of dermatoscopy—A new practical method for early recognition of malignant-melanoma,” *European Journal of Dermatology*, vol. 4, no. 7, pp. 521–527, 1994.
 - [18] A. G. Isasi, B. G. Zapirain, and A. M. Zorrilla, “Melanomas non-invasive diagnosis application based on the ABCD rule and pattern recognition image processing algorithms,” *Computers in Biology and Medicine*, vol. 41, no. 9, pp. 742–755, 2011.
 - [19] P. M. Roth and M. Winter, “Survey of appearance-based methods for object recognition,” Tech. Rep. ICG-TR-01/08, Institute of Computer Graphics and Vision, Graz University of Technology, Austria, 2008.
 - [20] K. Grauman and B. Leibe, “Visual object recognition,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 4, no. 2, pp. 1–181, 2011.
 - [21] M. Brown and D. Lowe, “Invariant features from interest point groups,” in *British Machine Vision Conference*, vol. 4, pp. 656–665, 2002.
 - [22] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
 - [23] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
 - [24] X. Jiang and J. Lai, “Sparse and dense hybrid representation via dictionary decomposition for face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, pp. 1067–1079, 2015.

- [25] X. Wang and X. Tang, “Dual-space linear discriminant analysis for face recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 564–569, IEEE, 2004.
- [26] S. J. D. Prince and J. H. Elder, “Probabilistic linear discriminant analysis for inferences about identity,” in *IEEE International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.
- [27] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti, “Tied factor analysis for face recognition across large pose differences,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 970–984, 2008.
- [28] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S. J. D. Prince, “Probabilistic models for inference about identity,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 144–157, 2012.
- [29] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.
- [30] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation,” in *Encyclopedia of Database Systems*, pp. 532–538, Springer, 2009.
- [31] R. Lowry, “Concepts and applications of inferential statistics,” 2014. [Online]. Available: <http://vassarstats.net/textbook/>.
- [32] S. G. Y. Ho and H. H. L. Chan, “The Asian dermatologic patient,” *American Journal of Clinical Dermatology*, vol. 10, no. 3, pp. 153–168, 2009.
- [33] W. Y. Lo and S. M. Puchalski, “Digital image processing,” *Veterinary Radiology & Ultrasound*, vol. 49, no. s1, 2008.
- [34] G. Paschos, “Perceptually uniform color spaces for color texture analysis: An empirical evaluation,” *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 932–937, 2001.
- [35] L. Shen and L. Bai, “A review on Gabor wavelets for face recognition,” *Pattern Analysis and Applications*, vol. 9, no. 2-3, pp. 273–292, 2006.

- [36] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, “CloudID: Trustworthy cloud-based and cross-enterprise biometric identification,” *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.
- [37] F. Bianconi and A. Fernández, “Evaluation of the effects of Gabor filter parameters on texture classification,” *Pattern Recognition*, vol. 40, no. 12, pp. 3325–3335, 2007.
- [38] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, IEEE, 2010.
- [39] S. Maji, A. C. Berg, and J. Malik, “Efficient classification for additive kernel SVMs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66–77, 2013.
- [40] A. Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms,” in *ACM International Conference on Multimedia*, pp. 1469–1472, ACM, 2010.