Interpretable Automatic Rosacea Detection with Whitened Cosine Similarity

1st Chengyu Yang
Department of Computer Science
New Jersey Institute of Technology
Newark, New Jersey, USA
cy322@njit.edu

2nd Chengjun Liu
Department of Computer Science
New Jersey Institute of Technology
Newark, New Jersey, USA
chengjun.liu@njit.edu

Abstract—According to the National Rosacea Society, approximately 16 million Americans suffer from rosacea, a common skin condition that causes flushing or long-term redness on a person's face. To increase rosacea awareness and to better assist physicians to make diagnosis on this disease, we propose an interpretable automatic rosacea detection method based on whitened cosine similarity in this paper. The contributions of the proposed methods are three-fold. First, the proposed method can automatically distinguish patients suffering from rosacea from people who are clean of this disease with a significantly higher accuracy than other methods in unseen test data, including both classical deep learning and statistical methods. Second, the proposed method addresses the interpretability issue by measuring the similarity between the test sample and the means of two classes, namely the rosacea class versus the normal class, which allows both medical professionals and patients to understand and trust the results. And finally, the proposed methods will not only help increase awareness of rosacea in the general population, but will also help remind patients who suffer from this disease of possible early treatment, as rosacea is more treatable in its early stages. The code and data are available at https://github.com/chengyuyang-njit/ICCRD-2025. The code and data are available at https://github.com/chengyuyang-njit/ICCRD-2025.

Index Terms—Statistical Learning, Deep Learning, Whitened Cosine Similarity, Computer-Aided Diagnosis, Medical Imaging, Explainability, Rosacea

I. INTRODUCTION

Rosacea is a commonly encountered chronic inflammatory skin disease in adults with a predilection for highly visible areas of the skin such as the face. It is characterized by flushing, redness, pimples, pustules and dilated blood vessels [1]. Rosacea symptoms often exhibit a cyclical pattern, with flare-ups lasting weeks to months before subsiding temporarily. In its early stages, however, rosacea may manifest as mild, transient facial redness (flushing), which can be easily overlooked. As a result, many patients may not medical attention until the condition has advanced, complicating treatment and management. Therefore, early detection is of extreme importance to facilitate timely intervention and improve clinical outcomes. Towards that end, we propose an interpretable automatic rosacea detection method based upon whitened cosine similarity.

The application of supervised deep learning methods to medical images, such as facial images of patients with rosacea, often encounters the challenge of insufficient labeled training data [2]. This issue is further complicated by the need to protect patient confidentiality, which limits data accessibility. Additionally, the rarity of certain diseases makes it even more difficult to gather an adequate amount of training data.

While research on rosacea and related skin conditions has been conducted using machine learning and computer vision/deep learning algorithms, most high-performing studies leveraging deep learning used datasets with nearly 10,000 images or even more [3][4][5][6]. However, the datasets in these works are fully confidential, making them difficult to reproduce [7].

As a result, deep learning methods tend to overfit the limited training data, which often leads to biased representation of the population distribution [8]. To solve the problem of lacking enough training data for rosacea, research has been conducted to generate more rosacea patients' frontal facial images [9] leveraging the generative adversarial networks (GAN) [10].

To simulate the real situation, that is, the amount of training data is limited and hard to collect, and also with the purpose of protecting patients' privacy, these generated rosacea patients' frontal facial images have been used to test the performance of some deep learning and statistical learning method [8]. According to the study, the reported accuracy of both methods on the real unseen test data is only 89.5% when the models are trained on these generated images. More specifically, the ResNet-18 method's recall rate is only 0.58. Which means 42% patients who are actually suffering from rosacea could not be diagnosed due to deep learning method's problem of overfitting to the training data and lack of generalizability to the unseen data. Also, the dark-box nature of deep learning method makes it lack of interpretability and less reliable from the perspective of medical professionals. Even though the statistical method using PCA proposed in the study achieves a relatively higher recall rate, the precision rate is only 0.75. If implemented in real medical environments, this method could generate a substantial number of false positives, potentially overwhelming clinicians with unnecessary alerts.

In this paper, we propose an interpretable automatic rosacea detection method leveraging the whitened cosine similarity to address the issues of both explainability and limitation of the training data. The contributions of our proposed method is three-fold. First, the proposed method effectively distinguishes

patients with rosacea from healthy individuals with significantly better performance than existing approaches in terms of the accuracy, recall rate and precision rate, etc. when tested on the unseen data. Second, it tackles the issue of interpretability by assessing the similarity between the test sample and the centroids of two distinct classes, rosacea and normal, allowing both medical practitioners and patients to easily understand and trust the outcomes. Finally, this method not only raises awareness of rosacea among the general population but also encourages timely early treatment, as rosacea is more manageable when diagnosed at earlier stages.

II. METHOD

We now present our automatic rosacea detection based upon whitened cosine similarity. It utilizes the means of the two classes, namely, the rosacea vs. the normal class, and measures the similarity between each test sample and the means of the two classes, giving an interpretable perspective for both the patients and the medical professionals.

Similarity measures play an important role in pattern recognition and computer vision [11]. Whitened Cosine Similarity is a refined variation of cosine similarity that addresses certain limitations in traditional cosine similarity by incorporating data whitening techniques. It is particularly useful in applications involving high-dimensional data, where redundant or correlated features can impact the performance of similarity measures. It has been robustly demonstrated for superior pattern recognition performance on large scale experiments [12]. It has also been shown that the whitened cosine similarity measure is related to the Bayes decision rule for minimum error under some specific assumptions [13].

Let δ_{WC} represent the whitened cosine similarity measure, which may be formulated as follows:

$$\delta_{WC}(\mathcal{U}, \mathcal{V}) = \frac{(W^t \mathcal{U})^t (W^t \mathcal{V})}{\|W^t \mathcal{U}\| \|W^t \mathcal{V}\|} \quad (1)$$

where $\mathcal{U}, \mathcal{V} \in \mathbb{R}^d$ are two pattern vectors, $\|\cdot\|$ denotes the norm operator, and W is the whitening transformation matrix, which may be specified by means of the covariance matrix. The covariance matrix of all samples, $\Sigma = \mathcal{E}\{(\mathcal{X} - M_0)(\mathcal{X} - M_0)^t\}$, can be factorized using principal component analysis (PCA): $\Sigma = \Phi \Lambda \Phi^t$, where $\mathcal{E}(\cdot)$ is the expectation operator, M_0 is the grand mean vector, Φ is an orthogonal eigenvector matrix, and Λ a diagonal eigenvalue matrix. The transformation, $W = \Phi \Lambda^{-1/2}$ is called the whitening transformation [14]. Note that, under this particular whitening transformation, (1) becomes

$$\delta_{WC}(\mathcal{U}, \mathcal{V}) = \frac{\mathcal{U}^t \Sigma^{-1} \mathcal{V}}{\|W^t \mathcal{U}\| \|W^t \mathcal{V}\|} \quad (2)$$

which is the commonly used form of the whitened cosine similarity measure, but it is not widely used in practice due to the complication of computing the inverse of the covariance matrix.

Our method leverages this similarity measure to build an intuitive and interpretable automatic rosacea detector. For each

test sample, we then calculate the whitened cosine similarity between this test sample and the means of the two classes: normal faces and faces having rosacea. If the similarity between the test sample and the mean of normal faces is larger than that between the test sample and the mean of faces having rosacea, the test sample is classified to be normal, otherwise, it is classified to be rosacea. The detailed algorithm is presented in Algorithm 1.

Algorithm 1: Automatic Rosacea Detection with Whitened Cosine Similarity

---

Input: $\mathbf{X}_{normal} \in \mathbb{R}^{d \times n}, \mathbf{X}_{rosacea} \in \mathbb{R}^{d \times m}, \mathcal{X} \in \mathbb{R}^d$
Output: *True / False*
/* GET ONE MATRIX OF ALL SAMPLES */
$\mathbf{X} \in \mathbb{R}^{d \times (n+m)} \leftarrow [\mathbf{X}_{normal}, \mathbf{X}_{rosacea}]$

/* CLASS MEAN */
$M_{normal} \leftarrow (\sum_{i=1}^{n} \mathbf{X}[:,i]) / n$
$M_{rosacea} \leftarrow (\sum_{i=n+1}^{n+m} \mathbf{X}[:,i]) / m$

/* ALL SAMPLES' MEAN */
$M_0 \leftarrow (\sum_{i=1}^{n+m} \mathbf{X}[:,i]) / (n+m)$

/* CENTER THE TRAINING DATA */
for $i = 1$ **to** $(n+m)$ **do**
$\quad | \quad \mathbf{X}[:,i] \leftarrow \mathbf{X}[:,i] - M_0$
end

/* UNBIASED ESTIMATE OF COVARIANCE */
$\Sigma \leftarrow \mathbf{X}\mathbf{X}^t / (n+m-1)$

/* PCA ON COVARIANCE MATRIX */
$\Sigma = \Phi \Lambda \Phi^t$

/* WHITENING TRANSFORMATION MATRIX */
$W \leftarrow \Phi \Lambda^{-1/2}$

/* COMPARE AND CLASSIFY */
if $\delta_{WC}(\mathcal{X}, M_{normal}) < \delta_{WC}(\mathcal{X}, M_{rosacea})$ **then**
$\quad | \quad$ return True
else
$\quad | \quad$ return False
end

We then compared our method with the deep learning and statistical learning method used in [8] on the same dataset. Other than that, we also implemented dimension reduction and feature extraction techniques like principal component analysis (PCA), and KNN [15] with different distance metrics (L1, L2, cosine similarity) for the purpose of comparison. K-Nearest Neighbors (KNN) is a simple, intuitive, and non-parametric machine learning algorithm used for both classification and regression tasks. It relies on the principle of similarity between data points and does not involve explicit training of a model. The similarity between data points can be measured

by Manhattan distance (L1), which is given by

$$d_{\text{Manhattan}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i| \qquad (3)$$

Or Euclidean distance (L2), which is given by

$$d_{\text{Euclidean}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (4)$$

Or cosine similarity, which is given by

$$\text{Cosine Similarity} = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \qquad (5)$$

III. EXPERIMENT

In this section, we first introduce the dataset, then we disclose some technical details which are especially important and helpful when implementing the algorithm 1 under the circumstances when the dimension of the data is extremely high. In the end, we compare and analyze the performance of our method with that of other classic deep learning and statistical learning methods in terms of metrics like accuracy, precision rate and recall rate, etc.

A. Dataset

We use the same dataset as is used in [8], for the same purpose of simulating the real situation where the amount of training data is limited and hard to collect, and also to protect patients' privacy.

1) Training and Validation Dataset: We use generated images for both rosacea positive and negative cases. As is shown in Fig 1, the frontal face images with rosacea have been generated using the GAN [9][10]. We therefore use 300 of the generated rosacea frontal face images and split them into the training set with 250 images and the validation set with the remaining 50 images. For the rosacea negative cases, we use 600 frontal face images generated from the Style-GAN [16] and split them into a training set with 500 images and the validation set with the remaining 100 images.

2) Test Dataset: To assess the performance of the proposed automatic rosacea detection method on real images, we use 50 real frontal face images with rosacea from various sources including Kaggle, DermNet, and National Rosacea Society. These images are selected with a standard of having at least 200 x 200 pixel resolution. The face in each image is detected and aligned using the DeepFace [17] and cropped and resized to the same size with the training and the validation images which are $512 \times 512 \times 3$ for the purpose of testing. In addition, we include 150 real people's frontal faces not having rasacea from the CelebA Dataset[18] in our test data set.

B. Technical Details Disclosure

When implementing the algorithm, the dimension of the data determines the size of the empirical covariance matrix. If the dimension of the data is very high ($d \gg n + m$), for example $512 \times 512 \times 3$ in our case, then it will be extremely

time consuming and memory inefficient to perform PCA on the covariance matrix.

Instead of performing PCA on Σ, which is a matrix of dimension $d \times d$ given by $\mathbf{X}\mathbf{X}^t$ multiplied by a constant $\frac{1}{n+m-1}$ that we ignore for now, we perform PCA on $\mathbf{X}^t\mathbf{X}$, which only has a dimension of $(n + m) \times (n + m)$. Let v be an eigenvector of $\mathbf{X}^t\mathbf{X}$ with eigenvalue λ, then

$$\mathbf{X}^t\mathbf{X}v = \lambda v \qquad (6)$$

Left multiply both sides of (6) with matrix $\mathbf{X}$, we get

$$\mathbf{X}(\mathbf{X}^t\mathbf{X}v) = \mathbf{X}(\lambda v) \qquad (7)$$

After applying matrix product associativity, we get

$$(\mathbf{X}\mathbf{X}^t)\mathbf{X}v = \lambda(\mathbf{X}v) \qquad (8)$$

In other words, if v is an eigenvector of $\mathbf{X}^t\mathbf{X}$, then $\mathbf{X}v$ is an eigenvector of $\mathbf{X}\mathbf{X}^t$ with the same eigenvalue. Therefore, we obtain the PCA on the much larger original empirical covariance matrix. It should also be noted that if v is a normalized eigenvector of $\mathbf{X}^t\mathbf{X}$, $\mathbf{X}v$ is not necessarily a normalized eigenvector of $\mathbf{X}\mathbf{X}^t$.

C. Performance Evaluation and Comparison

1) Evaluation Metrics: In this section, we introduce the evaluation metrics that we are going to use. To be more specific, we use accuracy, recall rate, precision rate, and F1 score. For a single method, all these four metrics are the higher, the better. The definitions of these four metrics are given below, where TP stands for the number of true positive predictions, TN stands for the number of true negative predictions, FP stands for the number of false positive predictions and FN stands for the number of false negative predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (11)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (12)$$

2) Performance Analysis: As we can see from TABLE I and TABLE II, our method achieved 100% accuracy on the validation data and 99% accuracy on the test data, significantly outperformed other methods in terms of accuracy, recall rate, precision rate and the F1 score. In the confusion matrix of our method on test data shown in TABLE III, only two out of fifty patients who actually suffer from rosacea are not diagnosed by our method, leading to the highest recall rate of all the listed methods. A higher recall rate will not only help increase rosacea awareness in the general population but also help remind the patients who suffer from this disease of
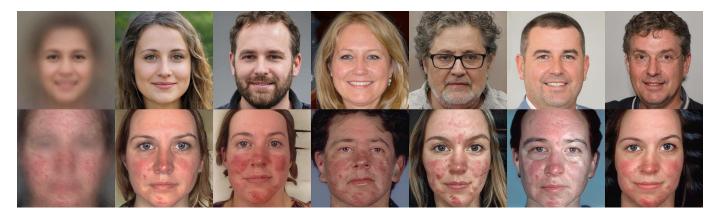
Fig. 1. The first column shows the mean images of the two training datasets from the rosacea negative and positive classes, respectively. The remaining six columns display six pair of random example images from the two training datasets corresponding to the normal people and those with rosacea, respectively.

TABLE I
DIFFERENT METHODS PERFORMANCE ON VALIDATION DATA

Method	Accuracy	Recall	Precision	F1
ResNet-18 [8][19]	1.00	1.00	1.00	1.00
Class dependent PCA [8]	0.90	0.96	0.77	0.86
Class independent PCA [14]	0.97	0.98	0.94	0.96
KNN with L1 metric [15]	0.99	1.00	0.98	0.99
KNN with L2 metric [15]	0.99	1.00	0.98	0.99
KNN with cosine metric [15]	1.00	1.00	1.00	1.00
KNN-L2 after PCA [15]	0.99	1.0	0.98	0.99
Our Method	**1.00**	**1.00**	**1.00**	**1.00**

TABLE II
DIFFERENT METHODS PERFORMANCE ON TEST DATA

Method	Accuracy	Recall	Precision	F1
ResNet-18 [8][19]	0.90	0.58	1.00	0.73
Class dependent PCA [8]	0.90	0.88	0.75	0.81
Class independent PCA [14]	0.93	0.84	0.875	0.86
KNN with L1 metric [15]	0.96	0.88	0.96	0.92
KNN with L2 metric [15]	0.95	0.88	0.90	0.89
KNN with cosine metric [15]	0.95	0.80	0.98	0.88
KNN-L2 after PCA [15]	0.95	0.82	0.95	0.88
Our Method	**0.99**	**0.96**	**1.00**	**0.98**

possible early treatment since rosacea is more treatable at its early stages.

It is also worth noting that, even though the classic deep learning method ResNet-18 achieves 100% accuracy on the validation set, due to the limitation on the size and the generated nature of the training and validation samples, it suffers from the problem of overfitting, thus cannot generalize well to the real world test dataset. Actually, TABLE II shows that its recall rate is only 0.58, which means 21 out of 50 rosacea positive cases are reported to be negative. In this case, a lot of patients actually having rosacea will not be diagnosed and will miss the best opportunity to start early treatment. Furthermore, deep neural networks' dark-box nature makes the result more unreliable from the perspective of medical practitioners.

3) Insights: Here we display the two rosacea test cases that our method misclassified as normal people's faces in Fig 2.

TABLE III
CONFUSION MATRIX OF OUR METHOD ON TEST DATA

	Classified as Normal	Classified as Rosacea
Actual Normal	150	0
Actual Rosacea	2	48

Fig. 2. The two test cases that our method misclassified.

As we can see, for the first case, the face area is too small compared to the face sizes in the training dataset. For the second case, the black area covering the patient's eyes is too large, which may extort the main features of the test image.

IV. CONCLUSION

In this paper, we propose a method for automatic rosacea detection utilizing whitened cosine similarity. demonstrates significant advancements in both accuracy and interpretability compared to existing approaches. By effectively distinguishing between rosacea and normal cases, our method achieves a high recall rate, ensuring that the majority of patients suffering from rosacea are accurately diagnosed. This is particularly crucial given the treatability of rosacea in its early stages, which can lead to improved patient outcomes. Furthermore, the interpretability of our approach allows both medical practitioners and patients to understand and trust the diagnostic results, fostering greater awareness of rosacea and encouraging timely intervention. As the challenges of limited training data and the dark-box nature of deep learning methods persist, our research highlights the importance of developing transparent and reliable diagnostic tools in the field of medical imaging. Future work will focus on further refining this method and

exploring its application in diverse clinical settings to enhance the management of rosacea and similar skin conditions.

REFERENCES

[1] Carsten Sauer Mikkelsen et al. "Rosacea: a clinical review". In: *Dermatology reports* 8.1 (2016).

[2] Tribikram Dhar et al. "Challenges of deep learning in medical image analysis—improving explainability and trust". In: *IEEE Transactions on Technology and Society* 4.1 (2023), pp. 68–75.

[3] Kenneth Thomsen et al. "Deep learning for diagnostic binary classification of multiple-lesion skin diseases". In: *Frontiers in medicine* 7 (2020), p. 574329.

[4] Zhixiang Zhao et al. "A novel convolutional neural network for the diagnosis and classification of rosacea: usability study". In: *JMIR medical informatics* 9.3 (2021), e23415.

[5] Haijing Wu et al. "A deep learning, image based approach for automated diagnosis for inflammatory skin diseases". In: *Annals of translational medicine* 8.9 (2020).

[6] Chen-Yu Zhu et al. "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment". In: *Frontiers in medicine* 8 (2021), p. 626369.

[7] Anwesha Mohanty et al. "Skin disease analysis with limited data in particular rosacea: a review and recommended framework". In: *IEEE Access* 10 (2022), pp. 39045–39068.

[8] Chengyu Yang and Chengjun Liu. "Increasing Rosacea Awareness Among Population Using Deep Learning and Statistical Approaches". In: *arXiv preprint arXiv:2411.07074* (2024).

[9] Anwesha Mohanty et al. "Towards Synthetic Generation of Clinical Rosacea Images with GAN Models". In: *2022 33rd Irish Signals and Systems Conference (ISSC)*. IEEE. 2022, pp. 1–5.

[10] Ian Goodfellow et al. "Generative adversarial networks". In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

[11] Kevin Bowyer and P Jonathon Phillips. *Empirical evaluation techniques in computer vision*. IEEE Computer Society Press, 1998.

[12] Hyeonjoon Moon and P Jonathon Phillips. "Computational and performance aspects of PCA-based face-recognition algorithms". In: *Perception* 30.3 (2001), pp. 303–321.

[13] Chengjun Liu. "The Bayes decision rule induced similarity measures". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.6 (2007), pp. 1086–1090.

[14] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Elsevier, 2013.

[15] Evelyn Fix. *Discriminatory analysis: nonparametric discrimination, consistency properties*. Vol. 1. USAF school of Aviation Medicine, 1985.

[16] Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4401–4410.

[17] Yaniv Taigman et al. "Deepface: Closing the gap to human-level performance in face verification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1701–1708.

[18] Ziwei Liu et al. "Deep learning face attributes in the wild". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 3730–3738.

[19] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.