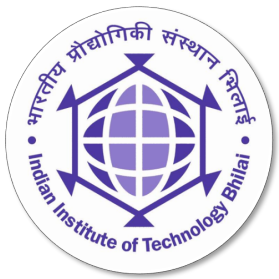
Improved fairness and accuracy of skin disease image classification using alignment with clinical text representations

Dr. Gagan Raj Gupta, Associate Professor, CSE, IIT Bhilai
Joint work with Aayushman, Hemant, Vidhi Mittal, Manisha Chawla at IIT Bhilai

June 22, 2025



The New York Times

Accou

Dermatology Has a Problem With Skin Color

Common conditions often manifest differently on dark skin. Yet physicians are trained mostly to diagnose them on white skin.

CONTRIBUTE

STAT

TOPICS OPINION TEAM EVENTS NEWSLETTERS REPORTS VIDEO PODCA

TRENDING: CORONAVIRUS FIRST OPINION BIOTECH HEALTH

HEALTH

Dermatology faces a reckoning: Lack of darker skin in textbooks and journals harms care for patients of color

By Usha Lee McFarling July 21, 2020 Reprints

Introduction

1. We introduce a novel approach to enhancing skin disease image classification accuracy and fairness through alignment with clinical text representations. Leveraging **Graph Optimal Transport (GOT)**, the proposed method aligns image-based representations with clinical textual representations.
2. The system addresses inherent biases and disparities in skin disease diagnosis through alignment with clinical text representations, promoting fairness and equity in healthcare outcomes.
3. We compare our model to SOTA model **FairDisCo** on two skin lesion datasets with different skin types: Fitzpatrick17k and Diverse Dermatology Images (DDI).
4. We adopt three fairness-based metrics **PQD**, **DPM** and **EOM** for our multiple classes and sensitive attributes task, highlighting the skin-type bias in skin lesion classification.

TABLE 1 Fitzpatrick Classification of Skin Types I through VI

Type I	Type II	Type III	Type IV	Type V	Type VI
White skin. Always burns, never tans.	Fair skin. Always burns, tans with difficulty.	Average skin color. Sometimes mild burn, tan about average.	Light-brown skin. Rarely burns. Tans easily.	Brown skin. Never burns. Tans very easily.	Black skin. Heavily pigmented. Never burns, tans very easily.
					

Figure: Fitzpatrick Scale (From Chapter 6: Clinical Presentation and Staging of Melanoma, William H. Ward et al.)

Dataset

	Skin Condition	Skin Type						
		T1	T2	T3	T4	T5	T6	Total
Fitz	Benign	444	671	475	367	159	44	2160
	Malignant	453	742	456	301	147	61	2160
	Non-neoplastic	2050	3395	2377	2113	1227	530	11692
	Total	2947	4808	3308	2781	1533	635	16012
DDI		T12		T34		T45		Total
	Malignant	49		74		48		171
	Non-malignant	159		167		159		485
	Total	208		241		207		656

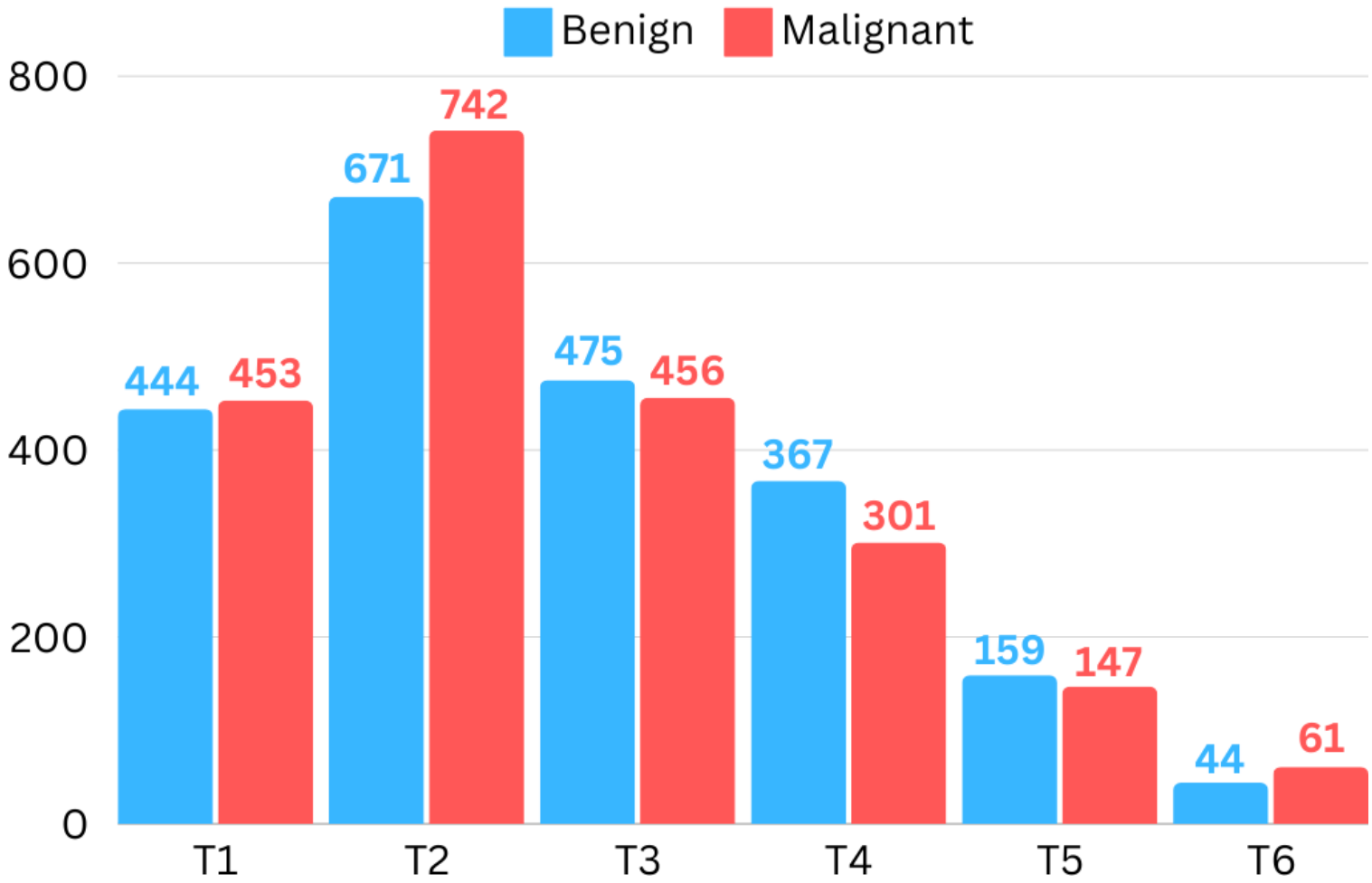


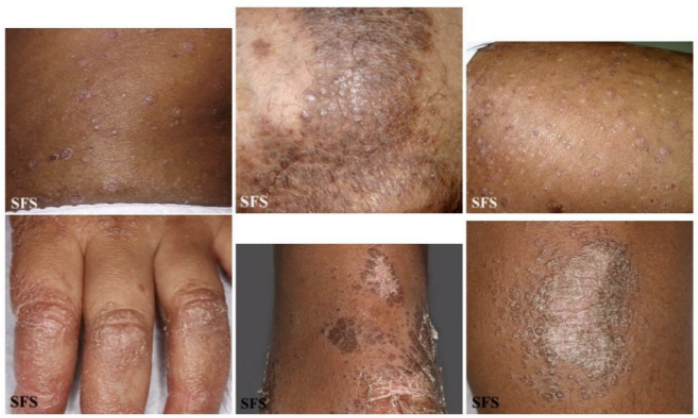
Figure: Class-Label Distribution of Benign and Malignant classes across skin tones in FitzPatrick17k Dataset

Fitzpatrick 17k

16,577 clinical images labeled with skin conditions and Fitzpatrick skin types

12,672 images from DermaAmin and 3,905 images from Atlas Dermatologico

	A	B	C	D	E	F	G
1	md5hash	fitzpatrick	label	nine_partition_label	three_partition_label	qc	url
2	5e82a45bc5d78bd24ae9202d194423f8	3	drug induced pigmentary changes	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/m/minocycline-pigmentation/minocycline-pigmentation1.jpg
3	fa2911a9b13b6f8af79cb700937cc14f	1	photodermatoses	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/p/photosensitivity/photosensitivity18.jpg
4	d2bac3c9e4499032ca8e9b07c7d3bc40	2	dermatofibroma	benign dermal	benign		https://www.dermaamin.com/site/images/clinical-pic/d/dermatofibroma/dermatofibroma71.jpg
5	0a94359e7eaacd7178e06b2823777789	1	psoriasis	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/p/psoriasis/psoriasis38.jpg
6	a39ec3b1f22				tic		https://www.dermaamin.com/site/images/clinical-pic/p/psoriasis-scalp/psoriasis-scalp20.jpg
7	45f7fe0e102						https://www.dermaamin.com/site/images/clinical-pic/k/kaposis-sarcoma/kaposis-sarcoma4.jpg
8	6c395be9325				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sweet-syndrome/sweet-syndrome98.jpg
9	9dc73230c77				tic		https://www.dermaamin.com/site/images/clinical-pic/g/granuloma_annulare/granuloma_annulare41.jpg
10	f23937e86de				tic		https://www.dermaamin.com/site/images/clinical-pic/L/larva-migrans/larva-migrans88.jpg
11	09d46db958				tic		https://www.dermaamin.com/site/images/clinical-pic/a/allergic_contact_dermatitis/allergic_contact_dermatitis114.jpg
12	9bc21ae9502				tic		https://www.dermaamin.com/site/images/clinical-pic/n/necrobiois-lipoidica-diabeticorum/necrobiois-lipoidica-diabeticorum88.jpg
13	e702b1a7dc4				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sweet-syndrome/sweet-syndrome50.jpg
14	ddcad677b7b				tic		https://www.dermaamin.com/site/images/clinical-pic/h/hidradenitis_suppurativa/hidradenitis_suppurativa50.jpg
15	b87804452f6						https://www.dermaamin.com/site/images/clinical-pic/L/lmm/lmm6.jpg
16	d1fb87ee7ee				tic	1 Diagnostic	https://www.dermaamin.com/site/images/clinical-pic/a/acne_vulgaris/acne_vulgaris150.jpg
17	8438db40abc				tic		https://www.dermaamin.com/site/images/clinical-pic/n/necrobiois-lipoidica-diabeticorum/necrobiois-lipoidica-diabeticorum7.jpg
18	2d57e08861t				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sarcoidosis-of-the-skin-plaque-form/sarcoidosis-of-the-skin-plaque-form15.jpg
19	1e119546f5b				tic		https://www.dermaamin.com/site/images/clinical-pic/X/xeroderma-pigmentosum/xeroderma-pigmentosum13.jpg
20	4c3f795cf8et						https://www.dermaamin.com/site/images/clinical-pic/m/melanoma/melanoma17.jpg
21	99247c9fe48						https://www.dermaamin.com/site/images/clinical-pic/d/dermatofibroma/dermatofibroma13.jpg
22	b09233673fc						https://www.dermaamin.com/site/images/clinical-pic/a/actinic_keratoses/actinic_keratoses83.jpg
23	449d63bec3a				tic		https://www.dermaamin.com/site/images/clinical-pic/L/localized-scleroderma/localized-scleroderma3.jpg
24	7a06b6baf51				tic		https://www.dermaamin.com/site/images/clinical-pic/h/hidradenitis_suppurativa/hidradenitis_suppurativa18.jpg
25	fb9640a13e0						https://www.dermaamin.com/site/images/clinical-pic/s/syringoma/syringoma33.jpg



- In the Fitzpatrick17k dataset, Groh et al. compiled 16,577 clinical images with skin condition labels and annotated them with Fitzpatrick skin-type labels. There are 114 different skin conditions, and each one has at least 53 images.
- They further divided these skin conditions into two more advanced categories: 3 (malignant, non-neoplastic, benign) and 9. Fitzpatrick labeling system is a six-point scale initially developed for classifying sun reactivity of skin and adjusting clinical treatment according to skin phenotype.
- The samples in the dataset are labeled by 6 Fitzpatrick skin types and 1 unknown type. **In our experiments, we ignore all samples of the unknown skin type**.

Cont.

- The DDI dataset contains 656 images with diverse skin types and pathologically confirmed skin condition labels, including 78 detailed disease labels and malignant identification.
- Fitzpatrick-12, Fitzpatrick-34, and Fitzpatrick-56, each contains a pair of skin-type classes, i.e., {T1,T2}, {T3,T4} and {T5,T6}, respectively

Figure: Example images from the entire DDI dataset for all skin tones (G), FST (Fitzpatrick Skin Type) I–II (H), and FST V–VI (I). Photo Credit: DDI dataset, Stanford School of Medicine.

Problem Formulation

In a multi-class skin lesion classification (M classes), the model is required to output a skin condition prediction y based on an RGB skin image $\boldsymbol{X} \in \mathbb{R}^{H \times W \times 3}$. We treat the skin type as a sensitive attribute s, including N groups with diverse types. Our goal is to model $p(y|\boldsymbol{X})$ without being affected by s.

Model Architecture

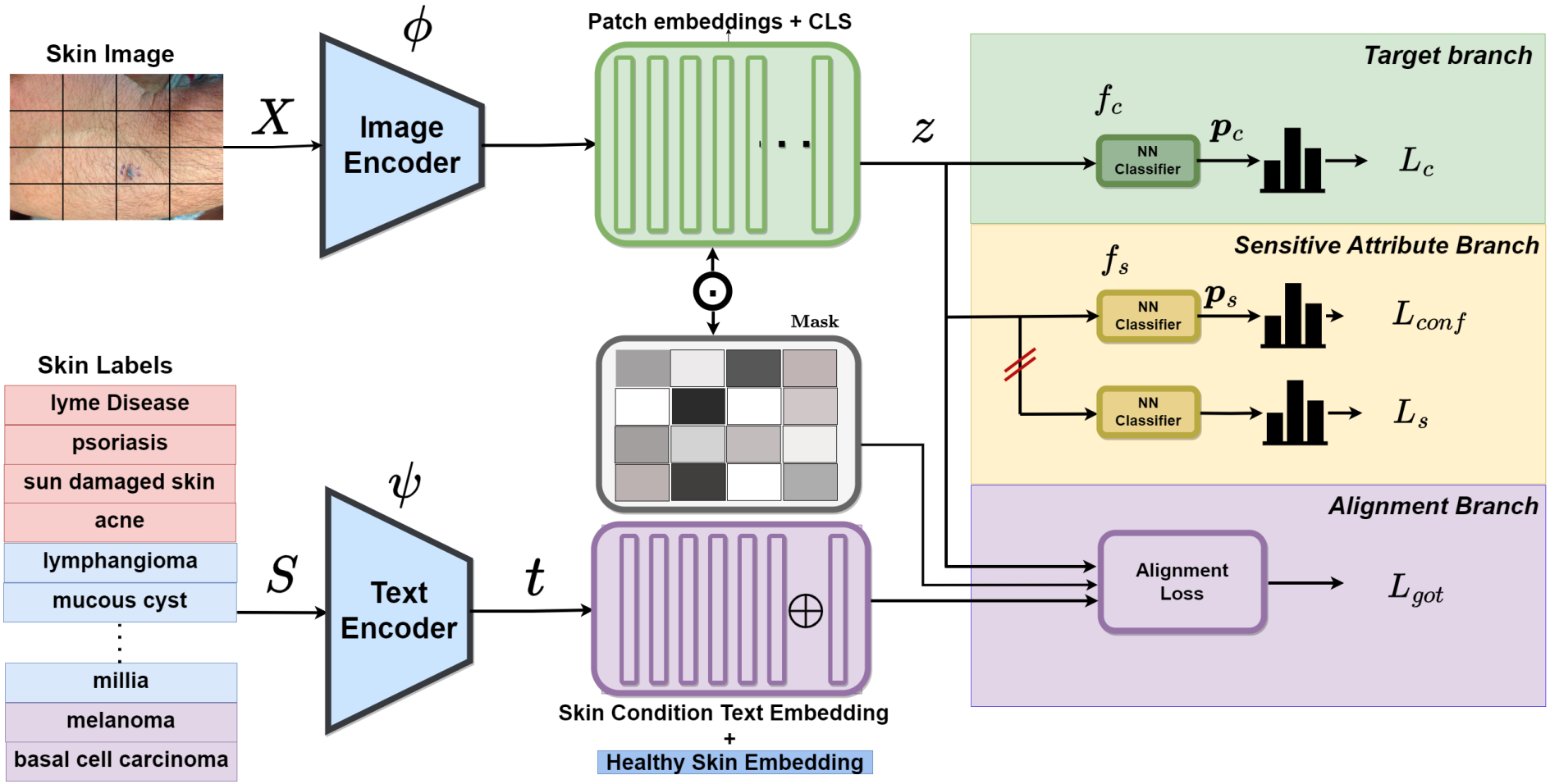


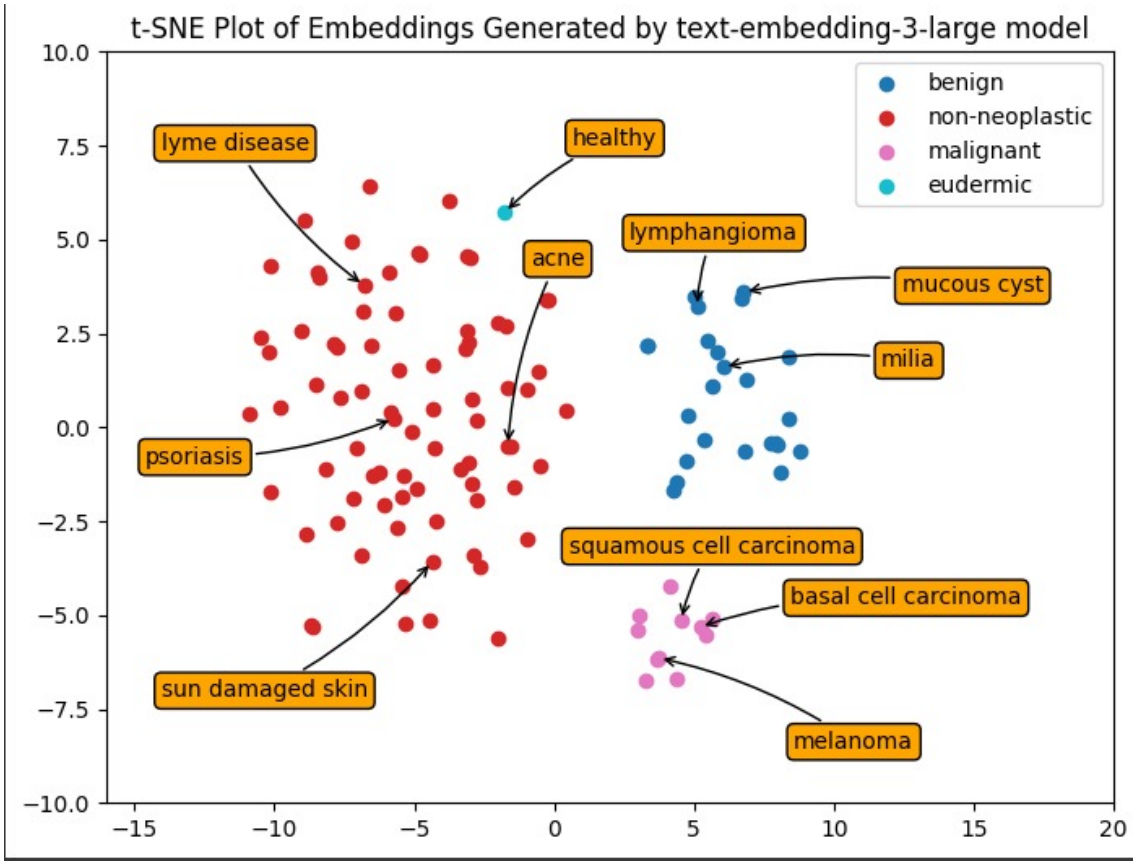
Figure: Model Architecture

Model Architecture

- Sensitive Attribute Branch is used to suppress skin tone features from text embedding.
- Alignment with Skin Disease Label to enhance features of patches. These labels have no bias of skin type.
- Eudermic (healthy) skin is used to align disease-free patches.
- Image Encoder converts the image into patches and produces patch representation with global image representation.

Embedding

We have used text embeddings to enhance the model by learning skin labels and the skin condition relationship as shown



Branch 1 - Classifier

Image is input to a feature extractor ϕ to get a representation $\boldsymbol{z} = \phi(\boldsymbol{X})$ that is then passed to three branches f_c, f_{sa} and f_{GOT}
Classifier includes a 3-layer feed-forward network and a softmax activation function, to get a skin condition prediction $\boldsymbol{p}^c$. We utilize a cross-entropy loss L_c on skin conditions to optimize the whole architecture.

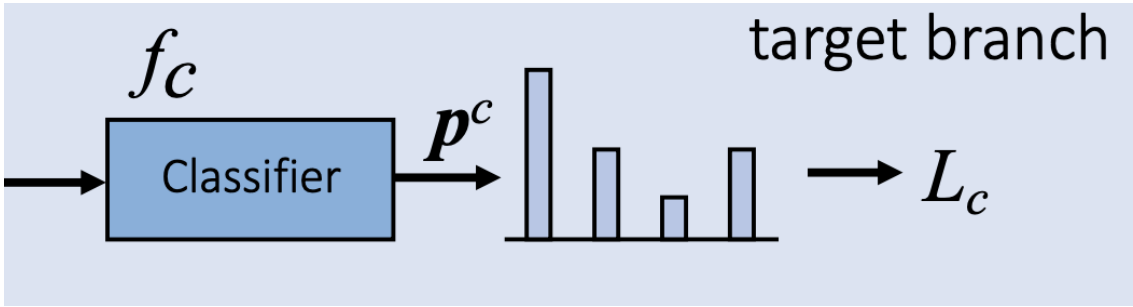


Figure: Classifier Branch

Branch 2 - Sensitive Attribute

- The ***SA*** branch consists of a classifier f_s to predict the skin type $\boldsymbol{p}_s$ based on the representation $\boldsymbol{z}$.
- We use adversarial cross-entropy loss for sensitive attributes on the SA branch and a gradient-reversal layer (GRL) between the feature extractor and the classifier.
- We minimize confusion loss given in Eq: 1 to confuse the feature extractor and remove the skin-type information from representations.

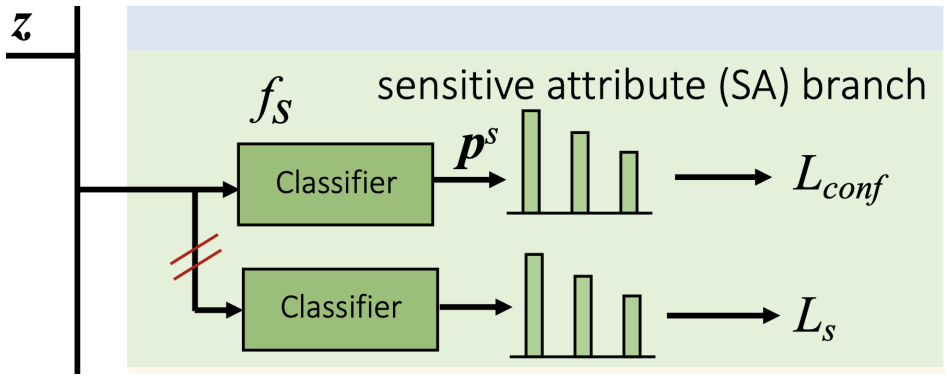


Figure: Sensitive Attribute Branch

$$L_{conf} = -\sum_{i=1}^{N} \frac{1}{N} \log(\boldsymbol{p}_i^s). \tag{1}$$

Branch-3 Alignment

- Image and text data inherently contain rich sequential/spatial structures. By representing them as graphs and performing graph alignment, not only can cross-domain relations be modeled, but also intra-domain relations are exploited.
- From Sub-label embedding and Image embedding from the data, we will construct a graph by calculating the similarity between a pair of elements in the embeddings.

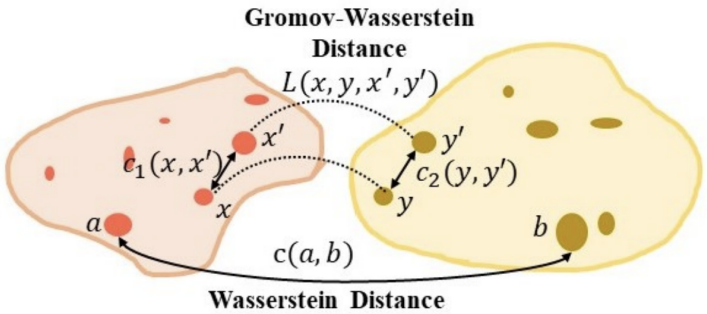


Figure: Calculation of WD and GWD from embeddings

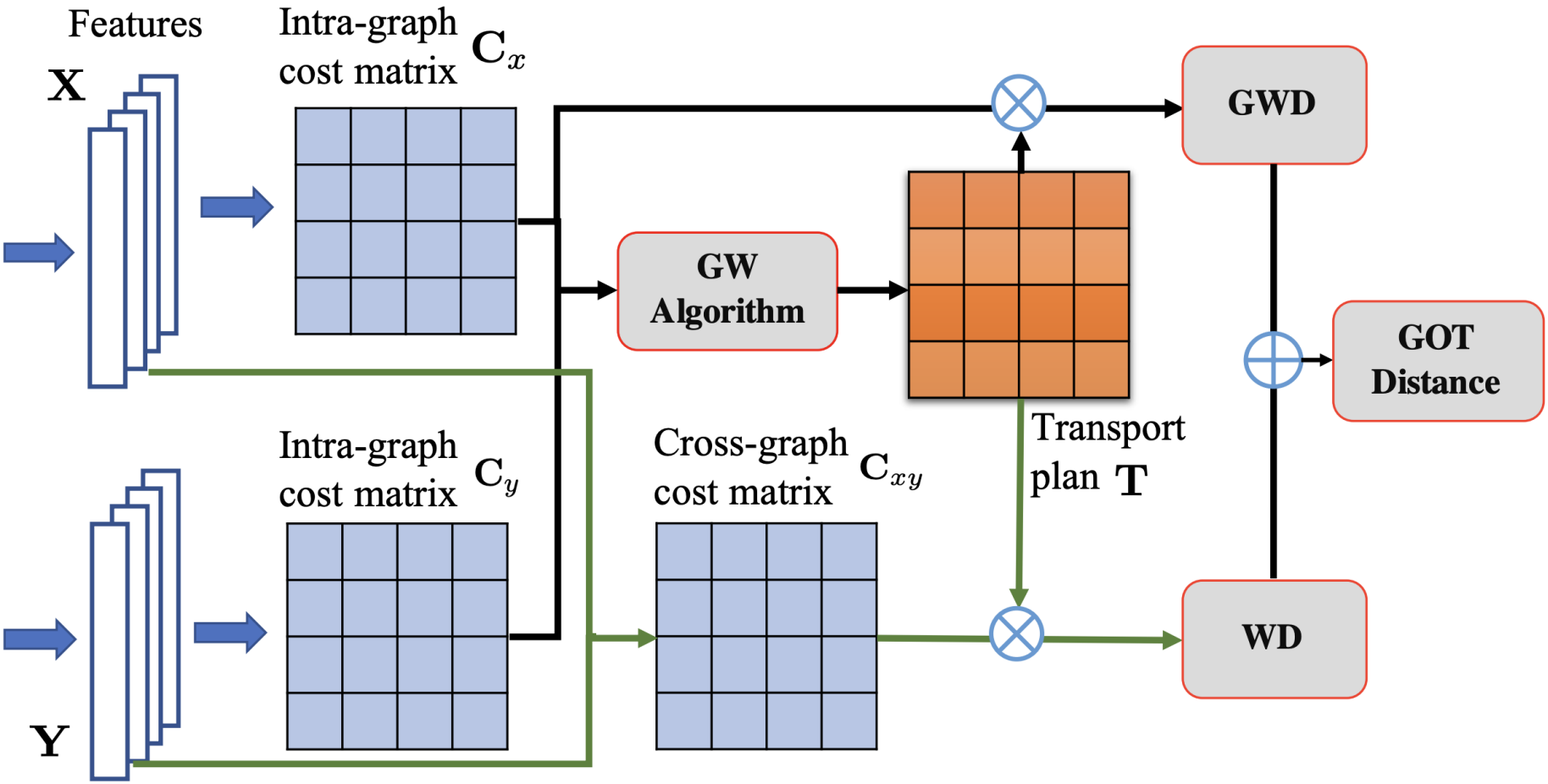


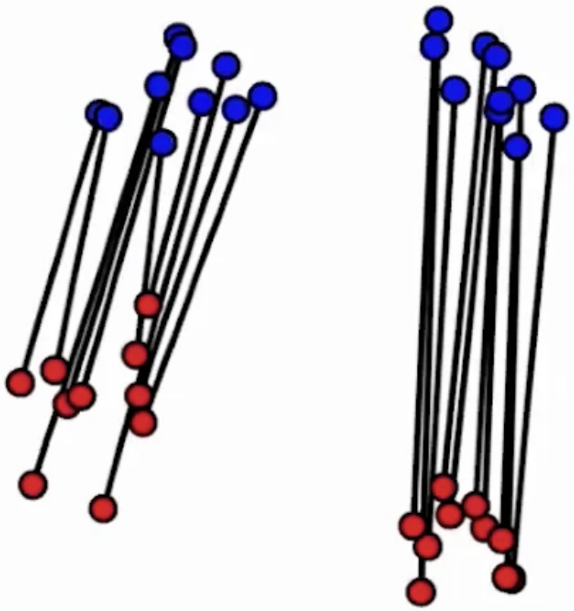
Figure: Calculation of GOT

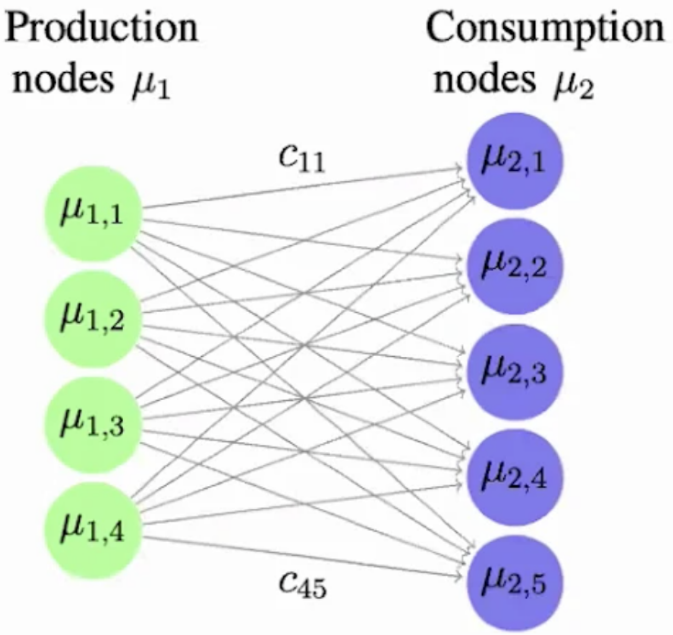
Discrete optimal transport

$$\min_{\mathbf{B} \in \mathbb{R}_+^{d_1 \times d_2}} \quad \text{trace}(\mathbf{C}^\top \mathbf{B})$$

$$\text{subject to } \mathbf{B}\mathbf{1} = \boldsymbol{\mu}_1$$

$$\mathbf{B}^\top \mathbf{1} = \boldsymbol{\mu}_2$$





Our Contribution

Learnable Mask

- We are proposing a novel learnable mask (M) on the transport plan that helps us in better alignment and fairness.

- MASK (M) can be used in $\mathcal{D}_{got}$ we put $\mathbf{T} \rightarrow M \odot \mathbf{T}$

- Extensive experiments, including both in-domain and out-domain evaluations on Fitzpatrick17k and DDI benchmarks.

Contribution Cont.

$$\mathcal{L}_{mw}(M, a, b) = \min_{\mathbf{T} \in \Pi(M,a,b)} \langle M \odot T, C \rangle$$

where
$\Pi(M, a, b) := \{\mathbf{T} \in \mathbb{R}^{n \times m} \mid$
$(\mathrm{M} \odot T)\mathbb{1}_m = a,\ (M \odot \mathbf{T})^T \mathbb{1}_n = b,\ \mathbf{T} \odot (\mathbb{1}_{n \times m} - M) = 0_{n \times m}\}.$

- We generate the Mask weights from patch embeddings using a generator network with a sigmoid activation function so that $M_{ij} \in [0, 1]$.
- The generated weights from the mask are then multiplied to patch embeddings. Thus, patches that align better with skin condition representations have higher weights than noisy patches.

Alignment Cont.

$$L_{got} = \mathcal{D}_{got}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\mathbf{T} \in \Pi(\mathbf{u}, \mathbf{v})} \sum_{i,i',j,j'} \mathbf{T}_{ij} \lambda c(\boldsymbol{x}_i, \boldsymbol{y}_j) + (1-\lambda) \mathbf{T}_{i'j'} c(\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{x}'_i, \boldsymbol{y}'_j) \,.$$

where $\Pi(\mathbf{u}, \mathbf{v})$ is the set of all transport plan
We apply the Sinkhorn Algorithm with an entropic regularizer to solve the above problem.

- We aim to backpropagate this loss to the image encoder for learning better representation of images.

$$L_{total} = L_c(\theta_\phi, \theta_{f_c}) + \alpha L_{conf}(\theta_\phi, \theta_{f_{sa}}) + L_s(\theta_{f_{sa}}) + \beta L_{got}(\theta_\phi, \theta_{label}, \texttt{Mask}) \quad (2)$$

Fairness Metrics

$$\text{Predictive Quality Disparity (PQD)} = \frac{\min(acc_j, j \in S)}{\max(acc_j, j \in S)} \tag{3}$$

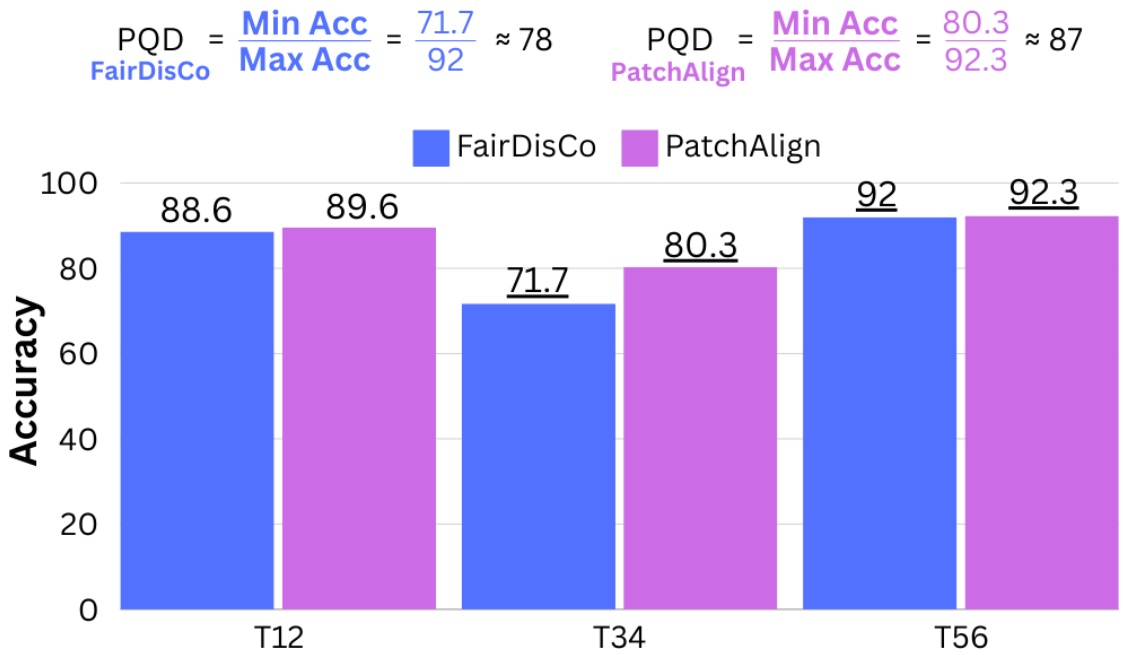


Figure: PQD calculated from DDI dataset across 3 skin types.

PQD measures the consistency in model accuracy across different skin types.

Equality of Opportunity Measure (EOM) =

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\min[p(\hat{y}=i|y=i,s=j),j\in S]}{\max[p(\hat{y}=i|y=i,s=j),j\in S]} \quad (4)$$

EOM evaluates fairness by comparing the true positive rates across groups, aiming for equal opportunity regardless of skin type. where S is set of skin type and Mn is number of labels

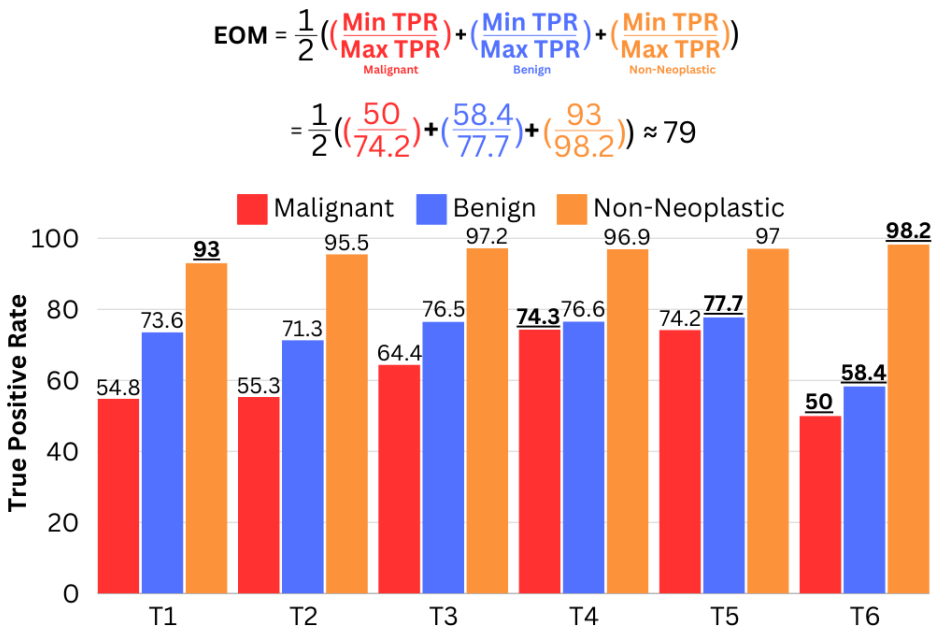


Figure: EOM calculated from Fitzpatrick17k dataset across 6 skin types, for PatchAlign.

Demographic Disparity Measure (DPM) =

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\min[p(\hat{y}=i|s=j), j\in S]}{\max[p(\hat{y}=i|s=j), j\in S]} \quad (5)$$

DPM evaluates how evenly the model's predictions are distributed among different demographic groups.

Table: In-Domain classification on Fitzpatrick17k Dataset

Model	Accuracy (%) ↑							Fairness Metrics (%)		
	Avg	T1	T2	T3	T4	T5	T6	PQD	DPM	EOM
BASE(ResNet-18)	85.0	82.6	81.7	85.0	88.8	90.7	82.8	88.9	52.5	64.0
RESM	85.1	82.5	81.8	86.2	89.0	91.2	82.3	89.0	50.7	62.0
REWT	85.6	83.4	83.0	86.2	89.0	90.6	83.5	90.8	50.0	62.9
ATRB	84.9	82.4	82.0	85.4	88.9	90.4	84.0	90.0	46.7	60.4
FairDisCo	85.1	82.2	82.1	86.2	89.4	90.0	83.2	90.2	51.2	63.8
MTL	88.3	**86.0**	**86.3**	88.9	**91.5**	91.3	90.0	**93.7**	50.5	71.2
PatchAlign	**88.6**	82.3	85.6	89.2	91.4	**91.3**	**91.4**	90.0	55.5	**74.8**
BASE(ViT)	86.0	83.5	82.9	86.9	90.5	90.5	84.6	89.4	47.0	59.2
BASE(ViT)+GOT	86.6	81.8	84.2	88.4	91.4	91.3	83.3	89.4	47.0	59.2
FairDisCo$^{\#}$	87.0	83.2	84.7	89.1	90.8	92.6	88.4	89.8	48.3	68.1
PatchAlign$^{\oslash}$	87.4	84.1	84.7	**90.3**	90.7	90.0	84.2	92.7	**62.7**	72.6

Table: Out-Domain Classification on the Fitzpatrick17k Dataset

E	Model	Accuracy							Fairness Metric		
		Avg	T1	T2	T3	T4	T5	T6	PQD	DPM	EOM
A	FairDisco	79.5	-	-	80.2	79.4	78.1	79.0	96.9	**73.5**	71.2
	PatchAlign	**84.0**	-	-	**84.2**	**84.0**	**83.1**	**84.6**	**98.3**	61.7	**68.2**
B	FairDisco	78.3	73.4	77.9	-	-	86.7	83.7	84.6	53.7	64.6
	PatchAlign	**82.4**	**78.6**	**82.1**	-	-	**89.5**	**86.1**	**87.8**	**70.7**	**75.1**
C	FairDisco	71.5	65.3	68.9	73.1	80.4	-	-	81.2	59.3	77.5
	PatchAlign	**77.6**	**74.4**	**75.6**	**78.2**	**83.5**	-	-	**89.1**	**74.7**	**78.1**

Links

- Scan to go to Github repository with code and link to the paper.

Why consider an entropic constraint in optimal transport?

The reason is computational,

Adding an entropy regularization to the optimal transport problem enforces a simple structure on the optimal regularized transport

Theorem 13.1 *The solution to entropic optimal transport,*

$$\begin{aligned} \min_{\pi} \quad & \sum_{ij} \pi_{ij} c_{ij} + \epsilon \sum_{ij} \log \pi_{ij} \\ s.t. \quad & \pi \mathbb{1} = a \\ & \pi^T \mathbb{1} = b \end{aligned}$$

is unique and has the form:

$$\pi^*_{ij} = u_i K_{ij} v_j \tag{1}$$

for scaling vectors $u \in \mathbb{R}^n, v \in \mathbb{R}^m$, *where* K *is a matrix with elements* $K_{ij} = \exp\left(-\frac{c_{ij}}{\epsilon}\right)$. *This means:*

$$\pi^* = Diag(u) K Diag(v). \tag{2}$$

This can be proven shown using Lagrange Multiplier Method.

Proof: The Lagrangian is the function:

$$L(\pi, \nu^{(1)}, \nu^{(2)}) = \sum_{ij} \pi_{ij} c_{ij} + \epsilon \sum_{ij} \pi_{ij} \log \pi_{ij} - {\nu^{(1)}}^T (\pi \mathbb{1} - a) - {\nu^{(2)}}^T (\pi^T \mathbb{1} - b) - \epsilon(\sum_{ij} \pi_{ij} - 1).$$

At optimality,

$$\begin{aligned} \frac{\partial L}{\partial \pi_{ij}} &= 0 \\ &= c_{ij} + \epsilon(\log \pi_{ij} + 1) - \nu_i^{(1)} - \nu_j^{(2)} - \epsilon \\ &= c_{ij} + \epsilon \log \pi_{ij} - \nu_i^{(1)} - \nu_j^{(2)} \\ \implies \log \pi_{ij}^* &= \frac{-c_{ij} + \nu_i^{(1)} + \nu_j^{(2)}}{\epsilon} \end{aligned}$$

$$\begin{aligned} \pi_{ij}^* &= \exp\left(\frac{-c_{ij} + \nu_i^{(1)} + \nu_j^{(2)}}{\epsilon}\right) \\ &= \exp\left(\frac{\nu_i^{(1)}}{\epsilon}\right) \exp\left(-\frac{c_{ij}}{\epsilon}\right) \exp\left(\frac{\nu_j^{(2)}}{\epsilon}\right). \end{aligned}$$

$$u = \exp\left(\frac{\nu_i^{(1)}}{\epsilon}\right)$$

$$v = \exp\left(\frac{\nu_j^{(2)}}{\epsilon}\right)$$

$$\text{s.t.} \quad \left[\text{Diag}(u) K \text{Diag}(v)\right] \mathbb{1} = a = \pi \mathbb{1}$$

$$\left[\text{Diag}(v) K^T \text{Diag}(u)\right] \mathbb{1} = b = \pi^T \mathbb{1}$$

Sinkhorn Algorithm

Thus, we get the following system of equations, with two equations and two unknowns:

$$\text{Diag}(u)Kv = a \tag{13.8}$$
$$\text{Diag}(v)K^T u = b. \tag{13.9}$$

To solve this system,

1. Initialize $v^{(0)} = (1, 1, \ldots, 1)^T$.

2. Update $u^{(1)}$ so that equation (13.8) holds.

3. Update $v^{(1)}$ so that equation (13.9) holds.

At the $(l+1)^{\text{th}}$ iteration, we would have used $v^{(l)}$ to update $u^{(l)}$, giving us $u^{(l+1)}$, which would then be used to update $v^{(l)}$, yielding $v^{(l+1)}$. This is shown below:

$$u^{(l+1)} = \frac{a}{Kv^{(l)}} \quad \rightarrow \quad v^{(l+1)} = \frac{b}{K^T u^{(l+1)}}. \tag{13.10}$$

For the n^{th} iteration, we would have:

$$\text{Diag}(u^{(n)}) \exp\left(-\frac{c_{ij}}{\epsilon}\right) \text{Diag}(v^{(n)}). \tag{13.11}$$

Multi-Task Approach for learning better representations

- We investigate multi-task learning (MTL) [**?**] to learn high-quality representations for skin cancer detection.
- MTL involves training a model on multiple tasks simultaneously. In our case, the two tasks are: 1) predicting meta-labels and 2) predicting the skin condition.
- The GOT loss is replaced with a combination of two losses: a predictive cross-entropy loss for the skin conditions and a contrastive loss for the skin labels used for classification, similar to FairDisCo [**?**].
- MTL implicitly encourages alignment, whereas PatchAlign explicitly addresses alignment through the GOT loss, potentially leading to more robust model performance.

Results and Discussion

Table: In-Domain classification on DDI Dataset

Model	Accuracy (%) ± Std-dev.				Fairness Metrics		
	Avg	T12	T34	T56	*PQD*	*DPM*	*EOM*
BASE	82.4 ± 1.5	83.3 ± 1.0	74.6 ± 5.7	89.7 ± 2.2	77.0 ± 1.9	75.2 ± 13.3	58.7 ± 4.3
FairDisCo	83.8 ± 0.4	88.6 ± 0.1	71.7 ± 2.2	92.0 ± 2.8	78.0 ± 4.5	72.8 ± 12	63.7 ± 3.5
MTL	82.3 ± 0.4	79.4 ± 3.3	82.6 ± 3.1	85.6 ± 1.5	91.4 ± 2.7	57.7 ± 5.9	77.2 ± 4.6
PatchAlign	87.4 ± 1.2	89.6 ± 2.6	80.3 ± 5.7	92.3 ± 1.3	86.9 ± 6.1	74.9 ± 12	69.6 ± 1.7