# BAB 2
# LANDASAN TEORI

## 2.1 Dataset DermaCon-IN

Bagian ini menjabarkan karakteristik dataset DermaCon-IN yang menjadi fondasi empiris seluruh eksperimen pada penelitian ini, dengan distribusi kelas dan tingkat *class imbalance* yang dirinci pada Tabel 2.1. Pemilihan dataset ini tidak bersifat kebetulan, sebab DermaCon-IN dikumpulkan dari populasi India Selatan yang secara demografis didominasi oleh *Fitzpatrick Skin Type* (FST) III–VI, yaitu rentang kulit sedang hingga gelap yang secara historis kurang terwakili pada dataset dermatologi konvensional yang didominasi populasi Barat berkulit terang [2, 7]. Karakteristik populasi inilah yang menjadikan DermaCon-IN sebagai konteks uji yang relevan untuk mengevaluasi apakah disparitas *fairness* demografis tetap muncul meskipun mayoritas populasi tidak didominasi oleh kulit terang, sebuah pertanyaan yang menjadi salah satu kontribusi utama penelitian ini.

### 2.1.1 Deskripsi dan Karakteristik Dataset

DermaCon-IN adalah dataset dermatologi klinis yang dikumpulkan secara prospektif dari klinik rawat jalan di India Selatan [1]. Dataset ini terdiri dari 5.450 citra yang berasal dari 3.002 pasien, dibagi menjadi 4.399 sampel *train* dan 1.051 sampel *test*. Citra-citra tersebut diklasifikasikan ke dalam 8 kelas penyakit berdasarkan kategori ICD-10, yakni: *Infectious Disorders*, *Inflammatory Disorders*, *Keratanisation Disorders*, *Neoplasms and Tumors*, *No Definite Diagnosis*, *Other Skin Disorders*, *Pigmentary Disorders*, dan *Skin Appendages Disorders*. Setiap citra dilengkapi dengan anotasi demografis berupa *Fitzpatrick Skin Type* (FST, skala I–VI) dan *Monk Skin Tone* (MST, skala 1–10), menjadikan dataset ini salah satu dari sedikit dataset dermatologi yang memungkinkan evaluasi *fairness* demografis secara granular [1].

### 2.1.2 Distribusi Kelas dan Class Imbalance

Distribusi kelas pada DermaCon-IN sangat tidak merata, sebagaimana ditunjukkan pada Tabel 2.1. Tabel tersebut merinci jumlah sampel pada setiap split *train* dan *test* untuk kedelapan kelas penyakit, beserta *imbalance ratio* (IR) yang

dihitung berdasarkan jumlah sampel pada split *train*, mengikuti Persamaan 2.1. Kolom Test ditampilkan secara terpisah untuk menunjukkan bahwa proporsi ketidakseimbangan pada data latih terjaga secara konsisten pada data uji, sehingga evaluasi performa model tetap dilakukan dalam kondisi distribusi kelas yang realistis, bukan pada data uji yang telah diseimbangkan secara artifisial.

Kelas mayoritas *Infectious Disorders* memiliki 1.808 sampel pada split *train* dan 419 sampel pada split *test*, sedangkan kelas minoritas *No Definite Diagnosis* hanya memiliki 31 sampel pada *train* dan 6 sampel pada *test*. Berdasarkan Persamaan 2.1, IR kelas *No Definite Diagnosis* terhadap kelas mayoritas dihitung sebagai IR $= 1.808/31 = 58{,}32\times$. Rentang IR yang sangat lebar ini, dari kelas kedua dengan IR $1{,}41\times$ hingga kelas ketujuh dengan IR $58{,}32\times$, menunjukkan bahwa ketidakseimbangan pada dataset ini tidak bersifat biner sederhana antara satu kelas mayoritas dan satu kelas minoritas, melainkan membentuk gradasi *long-tailed* di seluruh delapan kelas.

Tabel 2.1. Distribusi 8 kelas pada dataset DermaCon-IN. IR dihitung berdasarkan Persamaan 2.1, relatif terhadap kelas mayoritas pada split *train* (*Infectious Disorders*).

| Idx | Nama Kelas | Train | Test | IR (Train) |
|---|---|---|---|---|
| 0 | Infectious Disorders | 1.808 | 419 | 1,00 (mayoritas) |
| 1 | Inflammatory Disorders | 1.285 | 307 | 1,41 |
| 2 | Pigmentary Disorders | 667 | 164 | 2,71 |
| 3 | Skin Appendages Disorders | 379 | 101 | 4,77 |
| 4 | Other Skin Disorders | 131 | 28 | 13,80 |
| 5 | Keratanisation Disorders | 58 | 14 | 31,17 |
| 6 | Neoplasms and Tumors | 40 | 12 | 45,20 |
| 7 | No Definite Diagnosis | 31 | 6 | 58,32 (minoritas) |
| **Total** | | **4.399** | **1.051** | |

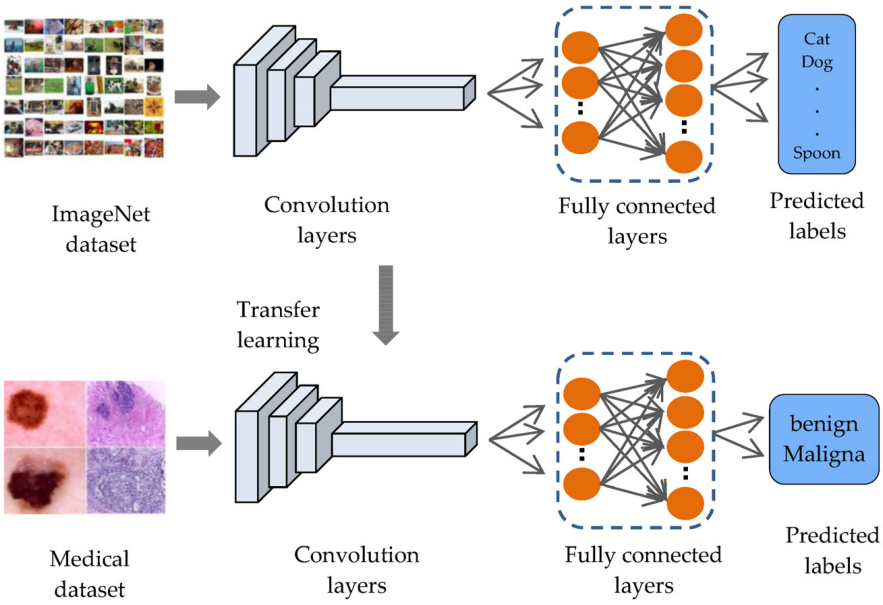
IR setinggi $58{,}32\times$ dikategorikan sebagai *extreme imbalance* [19]. Pada kondisi ini, model yang dilatih dengan *CrossEntropy Loss* standar cenderung memaksimalkan akurasi keseluruhan dengan cara mengabaikan kelas minoritas, sebuah fenomena yang disebut *majority class bias* [6].

## 2.2 Transfer Learning dan Vision Transformer

### 2.2.1 Transfer Learning

*Transfer learning* adalah paradigma pelatihan di mana bobot model yang telah dipra-latih (*pretrained*) pada dataset berskala besar digunakan sebagai

titik awal untuk melatih model pada tugas atau domain baru. Pendekatan ini sangat efektif ketika data pada domain target terbatas, karena bobot pra-latih sudah mengandung representasi fitur umum seperti tepi, tekstur, dan pola warna yang dapat ditransfer [20]. Dalam konteks klasifikasi penyakit kulit, *transfer learning* dari ImageNet secara konsisten menghasilkan performa yang lebih baik dibandingkan pelatihan dari awal (*from scratch*) [3].



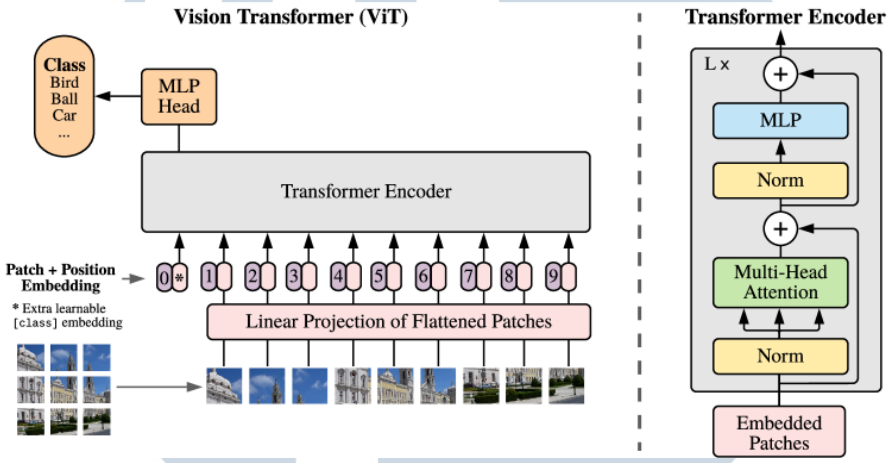

Gambar 2.1. *Transfer learning* dari model pretrained ImageNet ke domain target spesifik. *Sumber:* Abd El-Ghany et al. [21]

Gambar 2.1 mengilustrasikan konsep *transfer learning*: model yang telah dilatih pada dataset berskala besar (ImageNet-1K, 1.000 kelas objek umum) memiliki representasi fitur visual yang kaya dan dapat dimanfaatkan kembali untuk tugas domain spesifik dengan data yang lebih terbatas. Pada penelitian ini, bobot pretrained ImageNet-1K digunakan sebagai inisialisasi DaViT-Tiny sebelum *full fine-tuning* pada dataset DermaCon-IN (8 kelas penyakit kulit).

### 2.2.2 Vision Transformer (ViT)

*Vision Transformer* (ViT) [16] adalah arsitektur yang mengadaptasi mekanisme *self-attention* dari model Transformer NLP ke domain pengenalan gambar. Citra dibagi menjadi patch berukuran tetap (umumnya 16×16 piksel), setiap patch diproyeksikan menjadi vektor, dan sekuens vektor tersebut diproses oleh lapisan *Multi-Head Self-Attention* (MHSA). Berbeda dengan CNN yang

menangkap fitur lokal melalui konvolusi, ViT memiliki kemampuan *global receptive field* sejak lapisan pertama, sehingga mampu menangkap dependensi jarak jauh dalam citra. Kemampuan ini sangat relevan untuk penyakit kulit di mana pola distribusi global (misalnya pola simetri atau distribusi pigmen) menjadi indikator diagnostik penting.




Gambar 2.2. Arsitektur Vision Transformer (ViT).
*Sumber:* Dosovitskiy et al. [16]

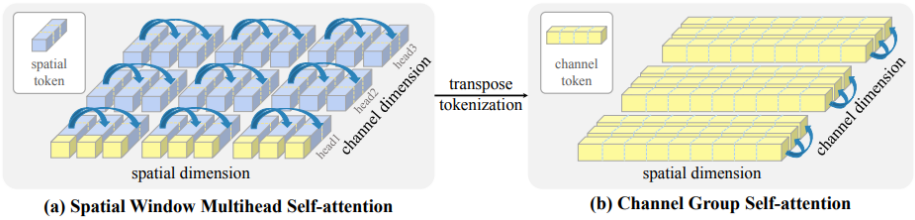
Gambar 2.2 menunjukkan arsitektur ViT secara keseluruhan. Citra masukan dipecah menjadi patch-patch berukuran tetap, kemudian diubah menjadi token melalui *linear projection of flattened patches*. Sebuah token `[class]` yang dapat dipelajari ditambahkan di awal urutan sebagai representasi global, bersama dengan *position embedding* untuk mempertahankan informasi posisi spasial. Seluruh urutan token kemudian diproses oleh *Transformer Encoder* yang terdiri dari lapisan *Multi-Head Attention*, *Norm*, dan MLP, dan token `[class]` pada keluaran digunakan oleh *MLP Head* untuk menghasilkan prediksi kelas [16].

### 2.2.3 Dual Attention Vision Transformer (DaViT)

DaViT (*Dual Attention Vision Transformer*) [15] adalah arsitektur hibrid yang menggabungkan dua jenis perhatian secara berurutan dalam setiap blok:

1. Spatial Window Self-Attention: Perhatian yang dihitung dalam jendela (*window*) lokal non-tumpang tindih untuk menangkap fitur spasial lokal yang efisien secara komputasi.

2. Channel Group Attention: Perhatian yang dihitung antar kanal pada seluruh *token* spasial untuk menangkap ketergantungan kanal global.

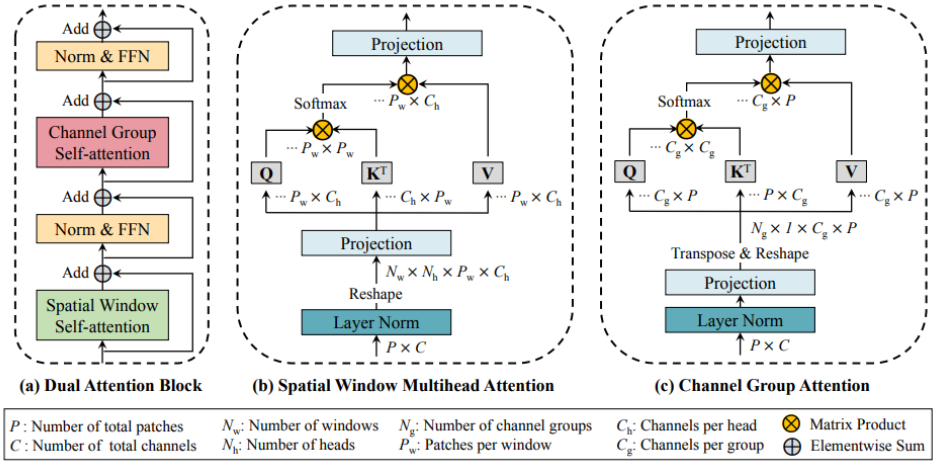



Gambar 2.3. Dua mekanisme *self-attention* pada DaViT: *Spatial Window* (kiri) dan *Channel Group* (kanan).

*Sumber:* Ding et al. [15]

Gambar 2.3 memperlihatkan dua mekanisme *self-attention* yang bekerja secara bergantian dalam DaViT. Pada *Spatial Window Multihead Self-attention* (kiri), setiap token hanya memperhatikan token lain dalam jendela spasial yang sama (*spatial token*), sehingga menangkap dependensi lokal secara efisien. Setelah *transpose tokenization*, token diorganisasi ulang sehingga *Channel Group Self-attention* (kanan) dapat beroperasi pada dimensi channel (*channel token*), menangkap dependensi global lintas posisi. Kedua mekanisme ini saling melengkapi untuk menghasilkan representasi fitur yang kaya secara spasial maupun semantik [15].

Dengan menggabungkan kedua mekanisme ini, DaViT mencapai performa yang kompetitif dengan kompleksitas komputasi yang lebih rendah dibandingkan ViT penuh. Varian DaViT-Tiny memiliki ≈28 juta parameter, sesuai untuk pelatihan pada GPU kelas konsumen seperti NVIDIA RTX 2060 Super (8 GB VRAM). Pada *benchmark* ImageNet-1K, DaViT-Tiny mencapai Top-1 accuracy 82,8%, mengungguli arsitektur serupa seperti Swin-T (81,3%) [15].




Gambar 2.4. *Dual Attention Block* pada DaViT beserta detail komputasi internal. *Sumber:* Ding et al. [15]

Gambar 2.4 merinci struktur internal *Dual Attention Block*. Setiap blok (a) terdiri atas dua sub-blok: *Spatial Window Self-attention* diikuti *Channel Group Self-attention*, keduanya dihubungkan melalui koneksi residual dengan lapisan Norm & FFN. Pada alur *Spatial Window Multihead Attention* (b), token diproyeksikan menjadi $\mathbf{Q}$, $\mathbf{K}^\top$, $\mathbf{V}$ berdimensi $P_w \times C_h$, lalu dikomputasi dengan Softmax. Pada alur *Channel Group Attention* (c), token di-*reshape* menjadi $C_g \times P$ sehingga attention beroperasi antar channel dalam satu grup [15].

## 2.3 Penanganan Class Imbalance

### 2.3.1 Definisi dan Dampak Class Imbalance

*Class imbalance* terjadi ketika distribusi label pada dataset pelatihan tidak merata [19]. Secara formal, untuk dataset dengan $C$ kelas, *imbalance ratio* didefinisikan sebagai:

$$\mathrm{IR} = \frac{n_{\max}}{n_{\min}} \tag{2.1}$$

di mana $n_{\max}$ adalah jumlah sampel kelas mayoritas dan $n_{\min}$ adalah jumlah sampel kelas minoritas. *Class imbalance* dikatakan *extreme* apabila IR $>$ 10 [19]. Dampak utamanya adalah model yang terlatih dengan *loss function* standar akan menghasilkan gradien yang didominasi oleh kelas mayoritas, sehingga kelas

minoritas terabaikan selama proses optimasi [6].

### 2.3.2 WeightedRandomSampler

*WeightedRandomSampler* adalah strategi penanganan *imbalance* berbasis *data-level*, di mana setiap sampel diberikan bobot sampling $w_i = 1/n_{c(i)}$ dengan $n_{c(i)}$ adalah jumlah sampel kelas dari sampel ke-$i$ [6]. Sampler ini bekerja pada level `DataLoader` PyTorch, yaitu dengan mengatur probabilitas terpilihnya setiap sampel dalam setiap batch sehingga ekspektasi distribusi kelas dalam setiap batch menjadi seragam, tanpa mengubah dataset asli maupun fungsi loss. Pendekatan ini setara secara statistik dengan *oversampling* kelas minoritas secara acak.

### 2.3.3 Class-Balanced Loss

*Class-Balanced Loss* (CB Loss) diusulkan oleh Cui et al. pada CVPR 2019 [22]. Metode ini mengintroduksi konsep *effective number of samples* untuk menggantikan jumlah sampel mentah sebagai dasar pembobotan:

$$E_n = \frac{1-\beta^n}{1-\beta} \tag{2.2}$$

di mana $n$ adalah jumlah sampel kelas dan $\beta \in [0,1)$ adalah hiperparameter yang mengontrol seberapa jauh sampel-sampel dalam kelas dianggap tumpang tindih (*overlapping*). Ketika $\beta \to 1$, $E_n \to n$ (pendekatan seperti frekuensi sampel). Ketika $\beta = 0$, $E_n = 1$ (semua kelas dianggap setara). Bobot kelas ke-$j$ kemudian dihitung sebagai:

$$w_j = \frac{1-\beta}{1-\beta^{n_j}} \tag{2.3}$$

Dalam penelitian ini digunakan $\beta = 0{,}9999$ sesuai rekomendasi penulis untuk dataset dengan IR tinggi [22]. CB Loss dengan $\beta = 0{,}9999$ menghasilkan bobot yang mendekati *inversely proportional* terhadap jumlah sampel, namun dengan regularisasi yang lebih halus dibandingkan *inverse frequency weighting* biasa.

### 2.3.4 Balanced Softmax Loss

*Balanced Softmax Loss* diusulkan oleh Ren et al. pada NeurIPS 2020 [23]. Motivasinya berangkat dari observasi bahwa saat inferensi, model seharusnya mempertimbangkan prior distribusi kelas $\pi_j = n_j / \sum_k n_k$. Metode ini memodifikasi logit dengan cara:

$$\tilde{z}_j = z_j + \log \pi_j \tag{2.4}$$

di mana $z_j$ adalah logit asli dari kelas ke-$j$ dan $\pi_j$ adalah prior frekuensi kelas. Koreksi ini diaplikasikan selama *training* sehingga keputusan batas (*decision boundary*) yang dipelajari model sudah memperhitungkan ketidakseimbangan distribusi kelas. Balanced Softmax secara implisit setara dengan *Logit Adjustment* dengan $\tau = 1$ (lihat Subbab berikut).

### 2.3.5 Logit Adjustment Loss

*Logit Adjustment Loss* diusulkan oleh Menon et al. pada ICLR 2021 [24]. Metode ini memberikan justifikasi teoritis yang lebih kuat dengan menunjukkan bahwa model yang dilatih dengan loss ini akan konvergen ke *Bayes-optimal classifier* untuk distribusi kelas yang sebenarnya. Formulanya adalah:

$$\tilde{z}_j = z_j + \tau \cdot \log \pi_j \tag{2.5}$$

di mana $\tau > 0$ adalah parameter yang dapat disetel untuk mengontrol kekuatan koreksi. Ketika $\tau = 1$, Logit Adjustment identik dengan Balanced Softmax. Perbedaan utama dengan Balanced Softmax adalah bahwa $\tau$ dapat di-*tune*, yaitu nilai $\tau > 1$ akan memberikan koreksi lebih kuat untuk kelas minoritas, sedangkan $\tau < 1$ melemahkan koreksi. Dalam penelitian ini digunakan $\tau = 1{,}0$ sebagai titik awal yang sesuai rekomendasi penulis [24].

### 2.3.6 Trade-Off antara Performa dan Fairness

Keempat strategi penanganan *class imbalance* di atas dirancang untuk mengoptimalkan metrik performa agregat seperti *macro F1* dan *balanced accuracy*, namun optimasi ini tidak secara otomatis menjamin distribusi perbaikan performa

yang merata di antara kelompok demografis. Secara teoritis, teknik *data-level* seperti *WeightedRandomSampler* mengoreksi ketidakseimbangan murni berdasarkan frekuensi label penyakit, tanpa memperhitungkan bagaimana frekuensi tersebut berinteraksi dengan distribusi warna kulit di dalam populasi [6]. Apabila kelas penyakit minoritas secara tidak proporsional terkonsentrasi pada kelompok warna kulit tertentu, maka penguatan sinyal gradien terhadap kelas tersebut berpotensi memperbesar, bukan mempersempit, kesenjangan akurasi antar kelompok FST maupun MST. Hal inilah yang mendasari mengapa penelitian ini tidak hanya mengevaluasi performa agregat, tetapi juga mewajibkan audit *fairness* demografis secara eksplisit pada setiap strategi yang diuji, sebagaimana dijabarkan pada Subbab 2.5.

## 2.4 Metrik Evaluasi Performa

### 2.4.1 Overall Accuracy

*Overall Accuracy* (OA) mengukur proporsi prediksi yang benar terhadap total prediksi:

$$\mathrm{OA} = \frac{\sum_{j=1}^{C} TP_j}{N} \tag{2.6}$$

di mana $TP_j$ adalah jumlah prediksi benar kelas ke-$j$ dan $N$ adalah total sampel. OA bersifat bias terhadap kelas mayoritas pada dataset *imbalanced*, sehingga tidak cukup dijadikan metrik utama dalam penelitian ini.

### 2.4.2 Balanced Accuracy

*Balanced Accuracy* (BalAcc) adalah rata-rata *recall* per kelas:

$$\mathrm{BalAcc} = \frac{1}{C} \sum_{j=1}^{C} \frac{TP_j}{TP_j + FN_j} \tag{2.7}$$

Metrik ini memberikan bobot yang sama pada setiap kelas, terlepas dari jumlah sampelnya, sehingga lebih sensitif terhadap performa pada kelas minoritas [25].

### 2.4.3 Macro F1-Score

*Macro F1* adalah rata-rata tidak berbobot dari F1-score setiap kelas:

$$\text{Macro F1} = \frac{1}{C}\sum_{j=1}^{C}\frac{2\cdot \text{Precision}_j \cdot \text{Recall}_j}{\text{Precision}_j + \text{Recall}_j} \tag{2.8}$$

Macro F1 menjadi metrik utama dalam penelitian ini karena memberikan gambaran yang seimbang antara *precision* dan *recall* di semua kelas, termasuk kelas minoritas.

## 2.5 Fairness dalam Machine Learning

### 2.5.1 Konsep Fairness

*Fairness* dalam *machine learning* mengacu pada properti bahwa sistem AI memberikan hasil yang adil dan setara kepada semua kelompok demografis [26]. Dalam konteks dermatologi komputasional, *fairness* berarti model harus memberikan performa diagnostik yang setara kepada pasien dengan warna kulit yang berbeda, tidak hanya pada populasi mayoritas di data pelatihan [2]. Ketidakadilan (*unfairness*) dapat muncul secara tidak langsung (*disparate impact*) meskipun model tidak secara eksplisit menggunakan atribut sensitif seperti warna kulit sebagai fitur input.

### 2.5.2 Fitzpatrick Skin Type (FST)

*Fitzpatrick Skin Type* (FST) adalah skala dermatologis yang membagi warna kulit manusia menjadi enam tipe berdasarkan respons terhadap paparan sinar matahari, mulai dari FST I (sangat terang, selalu terbakar) hingga FST VI (sangat gelap, tidak pernah terbakar) [9]. Skala ini banyak digunakan dalam penelitian *fairness* AI dermatologi karena merepresentasikan gradasi warna kulit yang relevan secara klinis [7]. Pada dataset DermaCon-IN, mayoritas sampel berada pada FST III hingga FST VI, mencerminkan demografi populasi India Selatan.

### 2.5.3 Monk Skin Tone (MST)

*Monk Skin Tone* (MST) adalah skala warna kulit yang dikembangkan oleh Ellis Monk dan Google [10] sebagai alternatif dari skala Fitzpatrick yang dianggap kurang merepresentasikan keragaman warna kulit populasi global. MST

terdiri dari 10 tingkatan (MST 1 = paling terang hingga MST 10 = paling gelap) dengan distribusi yang lebih merata pada spektrum gelap. Penggunaan MST bersamaan dengan FST dalam penelitian ini memberikan evaluasi *fairness* yang lebih komprehensif.

### 2.5.4 Metrik Fairness

Tiga metrik *fairness* digunakan dalam penelitian ini.

**Accuracy Gap** ($\Delta_{\text{ACC}}$) mengukur selisih akurasi tertinggi dan terendah di antara kelompok demografis:

$$\Delta_{\text{ACC}} = \max_{g \in \mathcal{G}} \text{Acc}(g) - \min_{g \in \mathcal{G}} \text{Acc}(g) \tag{2.9}$$

**Equalized Odds Gap** ($\Delta_{\text{EO}}$) [27] mengukur rata-rata selisih *True Positive Rate* (TPR) antar kelompok per kelas:

$$\Delta_{\text{EO}} = \frac{1}{C} \sum_{j=1}^{C} \left( \max_{g \in \mathcal{G}} \text{TPR}_j(g) - \min_{g \in \mathcal{G}} \text{TPR}_j(g) \right) \tag{2.10}$$

**Demographic Parity Gap** ($\Delta_{\text{DP}}$) [28] mengukur standar deviasi rata-rata *Positive Prediction Rate* (PPR) antar kelompok per kelas:

$$\Delta_{\text{DP}} = \frac{1}{C} \sum_{j=1}^{C} \text{std}_{g \in \mathcal{G}} \left( \text{PPR}_j(g) \right) \tag{2.11}$$

Untuk ketiga metrik di atas, nilai yang lebih rendah mengindikasikan model yang lebih adil (↓ = lebih adil). Target ideal yang ditetapkan dalam penelitian ini adalah $\Delta_{\text{FST\_ACC}} < 0{,}10$ dan $\Delta_{\text{MST\_ACC}} < 0{,}20$ [26].

### 2.5.5 Composite Score

Untuk mengintegrasikan dimensi performa dan *fairness* dalam satu nilai skalar, penelitian ini menggunakan *composite score* sebagaimana dirumuskan pada Persamaan 2.12. Rumus ini dipecah menjadi dua baris agar tetap terbaca dalam batas margin halaman, tanpa mengubah makna matematisnya:

$$S_{\text{comp}} = 0{,}35 \cdot \tilde{F}_1 + 0{,}25 \cdot \widetilde{\text{BalAcc}} + 0{,}10 \cdot \widetilde{\text{OA}} + 0{,}10 \cdot \tilde{R} + 0{,}20 \cdot \left(1 - \widetilde{\Delta}_{\text{FST}}\right) + 0{,}20 \cdot \left(1 - \widetilde{\Delta}_{\text{MST}}\right) \tag{2.12}$$

di mana simbol $\tilde{\cdot}$ menunjukkan nilai yang telah dinormalisasi dengan *min-max normalization*:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2.13}$$

Bobot 0,35 pada Macro F1 mencerminkan prioritas utama penelitian pada kemampuan mendeteksi semua kelas secara seimbang, sementara bobot gabungan 0,40 pada kedua komponen *fairness* (FST dan MST) mencerminkan pentingnya dimensi keadilan demografis yang menjadi kontribusi utama penelitian ini.

## 2.6 Penelitian Terkait

### 2.6.1 Klasifikasi Penyakit Kulit dengan Deep Learning

Esteva et al. [3] menunjukkan bahwa CNN berbasis Inception-v3 dapat mencapai performa setara dermatologis dalam mengklasifikasikan dua jenis kanker kulit dari 129.450 citra klinis. Studi ini menjadi tonggak yang membuktikan potensi *deep learning* dalam dermatologi komputasional. Liu et al. [4] selanjutnya mengembangkan sistem berbasis CNN untuk 26 kondisi kulit yang umum, mengungguli dokter umum namun masih di bawah dermatologis pada beberapa kondisi. Penelitian-penelitian ini umumnya menggunakan dataset dari populasi Barat dengan dominasi warna kulit terang, sehingga tidak dapat mencerminkan tantangan klinis pada populasi Asia Selatan.

### 2.6.2 Penanganan Class Imbalance pada Data Medis

Buda et al. [6] melakukan studi sistematik terhadap strategi penanganan *class imbalance* pada klasifikasi citra medis, menyimpulkan bahwa *oversampling* secara konsisten mengungguli *undersampling* dan bahwa kombinasi *oversampling* dengan *cost-sensitive learning* memberikan hasil terbaik. Cui et al. [22] memperkenalkan CB Loss yang menunjukkan peningkatan signifikan pada dataset

CIFAR, iNaturalist, dan Places. Ren et al. [23] mengusulkan Balanced Softmax yang memiliki justifikasi teoritis berbasis distribusi Boltzmann, sementara Menon et al. [24] memperkuat fondasi teoritis melalui kerangka *Bayes-optimal classification* dengan Logit Adjustment.

### 2.6.3 Fairness pada AI Dermatologi

Adamson & Smith [2] adalah studi pertama yang secara sistematis mendokumentasikan disparitas ras dan warna kulit dalam diagnosis dermatologi berbasis AI, menemukan bahwa model yang dilatih dengan data dominan FST I–III memiliki akurasi 10–15% lebih rendah pada pasien FST IV–VI. Groh et al. [7] melanjutkan dengan mengembangkan *benchmark* fairness untuk penyakit kulit menggunakan dataset Fitzpatrick17k, menunjukkan bahwa bias demografis persisten bahkan pada model mutakhir. Penelitian ini berkontribusi pada celah tersebut dengan menggunakan dataset populasi India Selatan yang memiliki distribusi FST berbeda dari dataset Barat, sekaligus mengevaluasi fairness dari perspektif teknik penanganan *class imbalance*.