Information System Journal (INFOS)
Vol. 8, No. 2, November 2025, pp. 215-224
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2025v8i02.2585

DETEKSI PENYAKIT KULIT DENGAN MENGGUNAKAN MODEL *PRETRAINED* DAN *HYBRID KNOWLEDGE DISTILLATION*

Theopilus Bayu Sasongko [1)], Arifiyanto Hadinegoro [2)], Eli Pujastuti [3)], I Made Artha Agastya [4)], Nazaruddin Ahmad [5)]

[1)2)3)] Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
[4)] Program Studi S2 Magister Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
[5)] Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry Banda Aceh

email: theopilus.27@amikom.ac.id [1)], arifiyanto@amikom.ac.id [2)], eli@amikom.ac.id [3)], artha.agastya@amikom.ac.id [4)], nazar.ahmad@ar-raniry.ac.id [5)]

INFO ARTIKEL

Riwayat Artikel:

Diterima November, 2025
Revisi November, 2025
Terbit November, 2025

ABSTRAK

Deteksi penyakit kulit berbasis citra medis membutuhkan model dengan akurasi tinggi namun tetap efisien untuk diterapkan pada perangkat dengan keterbatasan komputasi. Penelitian ini mengusulkan pendekatan *Hybrid Knowledge Distillation* (*Hybrid KD*) yang mengombinasikan distilasi berbasis logit (*soft target*), pembelajaran berbasis label asli (*hard label*), serta penyelarasan fitur internal (*feature alignment*) antara *teacher* dan *student model*. Model *teacher* dibangun menggunakan arsitektur *pretrained* berkapasitas besar, sedangkan model *student* menggunakan arsitektur ringan seperti *MobileNetV2*. Eksperimen dilakukan pada dataset citra penyakit kulit dengan pembagian data 80:10:10 untuk *training*, validasi, dan pengujian. Hasil evaluasi menunjukkan bahwa *Hybrid KD* secara konsisten meningkatkan performa student model dibandingkan metode distilasi konvensional. Secara khusus, *MobileNetV2* yang dilatih menggunakan *Hybrid KD* mencapai akurasi Top-1 sebesar 82.07%, melampaui skema KD berbasis logit maupun *feature alignment* secara terpisah, tanpa menambah kompleksitas model. Temuan ini menunjukkan bahwa *Hybrid KD* efektif dalam mentransfer pengetahuan *teacher* ke *student* dan berpotensi digunakan pada sistem diagnosis penyakit kulit berbasis perangkat dengan sumber daya terbatas.

Kata Kunci :
Knowledge Distillation; Penyakit Kulit; *Pretrained*; *Hybrid KD*

ABSTRACT

Image-based skin disease detection requires models that achieve high accuracy while remaining computationally efficient for deployment on resource-constrained devices. This study proposes a Hybrid Knowledge Distillation (Hybrid KD) approach that combines logit-based distillation (soft targets), hard-label supervision, and internal feature alignment between teacher and student models. The teacher model is constructed using a large capacity pretrained architecture, while the student model employs lightweight architecture such as MobileNetV2. Experiments are conducted on a skin disease image dataset with an 80:10:10 split for training, validation, and testing. The evaluation results demonstrate that Hybrid KD consistently improves the performance of the student model compared to conventional distillation methods. MobileNetV2 trained with Hybrid KD achieves a Top 1 accuracy of 82.07%, outperforming logit-based distillation and feature alignment applied independently, without increasing model complexity. These findings indicate that Hybrid KD effectively transfers knowledge from the teacher to the student and holds strong potential for deployment in skin disease diagnostic systems on resource-limited devices.

***Keywords*:**
Knowledge Distillation; *Skin Disease*; *Pretrained*; *Hybrid KD*

Penulis Korespondensi:

Arifiyanto Hadinegoro
Program Studi S1 Informatika,
Fakultas Ilmu Komputer,
Universitas Amikom Yogyakarta

Email:
arifiyanto@amikom.ac.id

1. PENDAHULUAN

Pembelajaran mendalam (*deep learning*) telah mencapai kemajuan yang sangat signifikan dalam bidang klasifikasi citra medis, khususnya pada bidang dermatologi [1], [2], di mana *Convolutional Neural Networks* (CNN) [3], [4] menunjukkan tingkat akurasi yang tinggi dalam mengidentifikasi dan mengklasifikasikan berbagai jenis penyakit kulit. Meskipun demikian, penerapan model-model tersebut di lingkungan klinis nyata masih menghadapi tantangan yang cukup besar, terutama akibat kebutuhan komputasi dan memori yang tinggi. Jaringan berkapasitas besar seperti *DarkNet-53* [5], *ResNet-152* [6], [7], dan *EfficientNet-B5* [8] mampu memberikan performa diagnostik yang sangat baik, namun memiliki biaya komputasi yang tinggi sehingga kurang sesuai untuk diterapkan pada perangkat dengan sumber daya terbatas, seperti aplikasi kesehatan berbasis *mobile* atau sistem klinis tertanam (*embedded systems*).

Keterbatasan ini menegaskan pentingnya pengembangan teknik kompresi model dan *transfer learning* yang mampu mempertahankan akurasi diagnostik sekaligus secara signifikan mengurangi beban komputasi. *Knowledge Distillation* (KD), yang pertama kali diperkenalkan oleh Hinton et al [9]., telah berkembang menjadi paradigma yang kuat untuk mentransfer pengetahuan dari model besar yang telah terlatih dengan baik (*teacher*) ke model yang lebih kecil dan ringan (*student*). Melalui pendekatan ini, waktu komputasi yang dibutuhkan untuk proses klasifikasi atau deteksi citra, video, maupun audio dapat dikurangi secara signifikan [10].

Berbagai penelitian sebelumnya telah mengeksplorasi pemanfaatan CNN dalam klasifikasi lesi kulit karena kemampuannya yang unggul dalam menangani tugas berbasis citra [11]. Hosny et al [12]., menggunakan *Deep Convolutional Neural Network* (DCNN) untuk mengidentifikasi *common nevus*, *atypical nevus*, dan *melanoma* pada dataset kanker kulit *PH2*. Dalam penelitian tersebut, arsitektur *AlexNet* dimanfaatkan untuk melakukan klasifikasi berbagai jenis keganasan kulit pada dataset *PH2*. *AlexNet* awalnya dirancang untuk tugas pengenalan visual pada *ImageNet*, dengan struktur yang terdiri dari lima lapisan konvolusi, satu lapisan *max pooling*, serta tiga lapisan *fully connected*. Pada penelitian ini, lapisan terakhir *AlexNet* digantikan dengan lapisan *softmax* untuk keperluan klasifikasi lesi kulit, dengan bobot model diperbarui menggunakan *stochastic gradient* dan disempurnakan melalui proses *backpropagation*.

Dalam konteks dermatologi, beberapa penelitian terkini mulai mengadopsi *transformer-based* model untuk klasifikasi lesi kulit. Flosdorf et al [13]., mengusulkan pendekatan *Vision Transformer* untuk klasifikasi *melanoma* pada dataset *ISIC*, yang menunjukkan peningkatan performa dibandingkan model *machine learning* tradisional seperti *KNN* dan *DT-Tree*, hasil menunjukan *Vit-L32* mendapatkan akurasi 91.57% dan *recall* 58.54%. Selain itu, Zhang et al [14]., mengembangkan arsitektur *DermVit* yang menggabungkan konteks *transformer* dengan *multi-scale context pyramid* yang dilatih pada *ISIC* 2018 dan *ISIC* 2019, hasil menunjukan performa 7.8% lebih unggul dibandingkan *ViT-Base*. Namun pendekatan berbasis *Vision Transformer* maupun *transfer learning* yang lain memiliki kekurangan di beban komputasi model yang relatif tinggi.

Dalam pendekatan KD, model *student* tidak hanya belajar dari label keras (*hard label*) yang tersedia pada dataset, tetapi juga dari distribusi probabilitas lunak (*soft label*) yang dihasilkan oleh model *teacher*. *Soft* target ini mengandung informasi yang lebih kaya mengenai kesamaan antar kelas serta hubungan antar kelas, sehingga memungkinkan model *student* memiliki kemampuan generalisasi yang lebih baik dibandingkan jika hanya dilatih menggunakan label *ground-truth*. Meskipun sederhana dan terbukti efektif, metode KD klasik masih menghadapi keterbatasan dalam mentransfer pengetahuan representasional yang mendalam serta fitur struktural, terutama pada tugas pencitraan medis yang kompleks, di mana granularitas fitur dan hubungan spasial memegang peranan penting. KD standar pada umumnya berfokus pada penyelarasan distribusi keluaran antara *teacher* dan *student*, namun sering kali mengabaikan korelasi fitur tingkat dalam serta panduan dari *hard label* yang sangat krusial dalam konteks diagnosis medis. Pada klasifikasi lesi kulit, misalnya, perbedaan tekstur yang halus, variasi warna, dan struktur morfologis merupakan faktor penting untuk membedakan antara kasus jinak dan ganas. Tidak adanya penyelarasan fitur secara langsung dapat menyebabkan hilangnya informasi spasial yang kritis pada model student. Oleh karena itu, pengembangan formulasi KD dengan memasukkan penyelarasan fitur dan penguatan *hard label* dipandang mampu memberikan transfer pengetahuan yang lebih komprehensif dan efektif. Dalam konteks tersebut, penelitian ini mengusulkan pengembangan kerangka *KD Hinton* dengan mengintegrasikan supervisi tingkat fitur (*feature-level supervision*) serta pembelajaran label hibrida untuk menjembatani kesenjangan representasi antara jaringan *teacher* dan *student*. Dengan adanya penyelarasan fitur, model student didorong untuk meniru representasi internal model *teacher*, sehingga pola spasial yang bersifat diskriminatif dapat dipertahankan dan tidak

Information System Journal (INFOS)
Vol. 8, No. 2, November 2025, pp. 215-224
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2025v8i02.2585

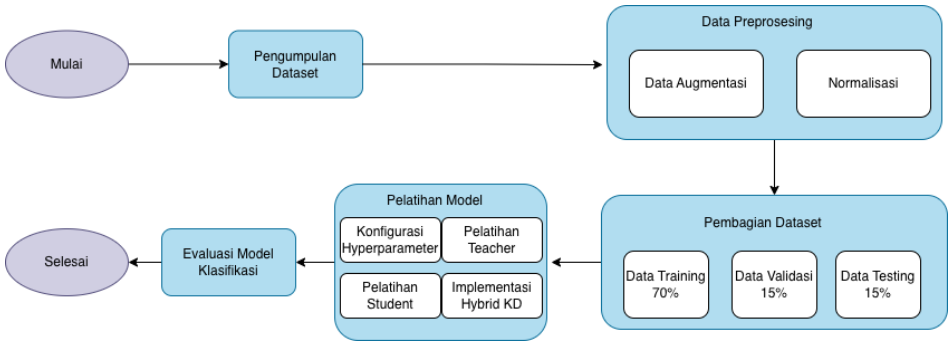
tereduksi selama proses distilasi. Di sisi lain, integrasi informasi *hard label* memastikan bahwa pengetahuan lunak yang ditransfer tidak mengaburkan batas kelas utama berdasarkan *ground truth*.

Melalui strategi distilasi hibrida ini, model *student* mampu mempelajari hubungan semantik tingkat tinggi sekaligus ciri diagnostik yang bersifat detail, yang sangat penting untuk mencapai performa klasifikasi yang andal pada domain citra medis yang kompleks seperti diagnosis lesi kulit. Pada penelitian ini, diusulkan tiga pengembangan terhadap kerangka *KD Hinton*, yaitu: *KD Hinton* dengan *Hard Label*, *KD Hinton* dengan *Feature Alignment*, dan *Hybrid KD Hinton*. Model *teacher* yang digunakan mencakup arsitektur berkapasitas tinggi seperti *DarkNet-53*, *ResNet-152*, *RepViT-M2-3* [15], *EfficientNet-B3/B5*, dan *WideResNet-50/101*, sedangkan model *student* terdiri dari *MobileNetV2* [16], *ResNet-8*, *ShuffleNetV2* [17], dan *EfficientNet-B0*.

Konfigurasi ini memungkinkan evaluasi mendalam terhadap kemampuan masing-masing varian distilasi dalam menyeimbangkan kinerja dan efisiensi komputasi pada tugas klasifikasi penyakit kulit. Dengan memanfaatkan model *teacher* pra-latih berkapasitas tinggi, penelitian ini mampu menangkap representasi fitur dermoskopik yang kaya, seperti pola pigmentasi, batas lesi, dan ketidakteraturan tekstur. Ketika ditransfer ke model *student* yang ringkas, representasi tersebut dapat meningkatkan presisi klasifikasi tanpa menambah beban komputasi secara signifikan. Selain itu, integrasi penyelarasan fitur dan mekanisme distilasi hibrida terbukti mampu mengurangi degradasi informasi selama proses kompresi model, sehingga model *student* tetap mempertahankan ciri diagnostik penting sekaligus mendukung inferensi yang cepat, yang sangat dibutuhkan pada sistem pendukung keputusan klinis secara *real-time*.

2. METODOLOGI PENELITIAN

Pada penelitian ini, diusulkan suatu model *knowledge distillation* yang ditingkatkan dengan mengintegrasikan beberapa pendekatan, yaitu distilasi berbasis hard label, penyelarasan fitur (feature alignment), serta kombinasi hibrida dari kedua pendekatan tersebut, guna meningkatkan kinerja identifikasi penyakit kulit. Berikut alur penelitian ada pada Gambar 1.



Gambar 1. Alur penelitian.

2.1 Pengumpulan Dataset

Dataset dikumpulkan dengan menggunakan data sekunder yang berasal dari dua publik dataset. Dataset penyakit kulit [18] (*ISIC* dataset dan *skin disease classification* dataset). Dataset ini memiliki 6 kelas label diantaranya adalah *Acne*, *Benign*, *Eczema*, *Infectious*, *Malignant*, dan *Pigment* yang terdiri dari 38753 gambar.

2.2 Normalisasi dan Data Augmentasi

Data citra dermoskopi pada dataset *ISIC* memiliki resolusi dan karakteristik visual yang beragam, sehingga pada tahap awal dilakukan penyeragaman resolusi input menjadi 116×116 piksel. Langkah ini bertujuan untuk menyeimbangkan beban komputasi serta memastikan konsistensi ukuran masukan pada seluruh model yang digunakan, terutama dalam skenario pelatihan dengan sumber daya komputasi terbatas. Meskipun resolusi relatif kecil, informasi tekstur dan pola visual yang relevan untuk diagnosis penyakit kulit tetap dapat dipertahankan pada skala tersebut. Selanjutnya, proses data augmentasi diterapkan untuk meningkatkan keberagaman data pelatihan dan memperbaiki kemampuan generalisasi model. Metode augmentasi yang digunakan meliputi *Random Horizontal Flip* [19], *Color Jitter* [20], dan *Random Rotation*, sebagaimana ditunjukkan pada Tabel 1. Strategi augmentasi ini dipilih secara konservatif dan terkontrol dengan

mempertimbangkan sifat klinis citra penyakit kulit. Transformasi geometrik ekstrem seperti rotasi besar, distorsi perspektif, atau *elastic deformation* tidak diterapkan karena berpotensi mengubah struktur morfologis lesi secara tidak realistis dan dapat menurunkan validitas klinis citra.

Tabel 1. Augmentasi data ISIC small.

No	Augmentasi	Value
1	*RandomHorizontalFlip*	*True*
2	*Color Jitter*	*brightness* = 0.3, *contrast* = 0.3, *saturation* = 0.3, *hue* = 0.05
3	*Random Rotation*	10
4	*Normalize*	*norm_mean* = (0.5, 0.5, 0.5), *norm_sd* = (0.5, 0.5, 0.5)

Selain itu, augmentasi berbasis warna seperti *Color Jitter* digunakan untuk mensimulasikan variasi kondisi pencahayaan, serta variasi warna kulit pasien, yang umum dijumpai dalam praktik klinis nyata. Augmentasi ringan namun bermakna secara klinis terbukti lebih efektif dibandingkan augmentasi agresif yang berisiko memperkenalkan artefak visual non-klinis. Dengan demikian, kombinasi augmentasi yang digunakan pada penelitian ini dirancang untuk menjaga keseimbangan antara peningkatan variasi data dan preservasi karakteristik medis yang esensial.

2.3 Pembagian Dataset

Pembagian Dataset (*Data splitting*) merupakan tahapan yang mengacu pada proses membagi dataset menjadi dua atau lebih bagian dimana biasanya dibagi menjadi dua yaitu data *training*, data validasi dan data *testing*. Data *training* digunakan untuk melakukan pembelajaran terhadap model *pretrained* dan *knowledge distillation* yang digunakan dalam melakukan klasifikasi gambar penyakit kulit. Data validasi digunakan untuk memvalidasi model sebelum dilakukan *test* pada data. Data *testing* dilakukan untuk melihat performa model dalam melakukan klasifikasi. Pada penelitian ini dilakukan pembagian dataset yaitu 80:10:10 seperti pada Tabel 2.

Tabel 2. Pembagian dataset.

Dataset	Jumlah Gambar
Training	30909
Validation	3922
Testing	3922

Selain pembagian data, penelitian ini juga menerapkan *class weight* [21] untuk mengatasi permasalahan ketidakseimbangan kelas (*class imbalance*) yang terdapat pada dataset. Penerapan *class weight* bertujuan untuk memberikan bobot yang lebih besar pada kelas dengan jumlah data yang lebih sedikit, sehingga kesalahan prediksi pada kelas minoritas memperoleh penalti yang lebih tinggi selama proses pelatihan. Dengan strategi ini, model diharapkan mampu belajar secara lebih adil terhadap seluruh kelas dan tidak bias terhadap kelas mayoritas, sehingga performa klasifikasi, khususnya pada kelas minoritas, dapat ditingkatkan.

2.4 Pelatihan Model

2.4.1. Konfigurasi Hyperparameter

Pelatihan yang ada menggunakan model *teacher* dan *student* yang beranekaragam. Sebelum dilakukan pelatihan digunakan beberapa *hyperparameter* seperti pada Tabel 3.

Tabel 3. *Hyperparameter settings*.

No	Hyperparameter	Value
1	*Epochs Teacher*	100
2	*Epochs Student*	100
3	*Batch Size*	32-96
4	*Resolutions*	116 x 116
5	*Learning Rate*	0.05
6	*Seed*	42
7	*Optimizer*	SGD
8	*Early Stopping*	100

Information System Journal (INFOS)
Vol. 8, No. 2, November 2025, pp. 215-224
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2025v8i02.2585

Penggunaan *SGD* dengan *learning rate* relatif tinggi [22], [23] terbukti efektif dalam meningkatkan kemampuan generalisasi model, khususnya ketika digunakan pada model *pretrained* yang dilatih ulang (*fine-tuning*) pada dataset medis seperti *ISIC*. Penelitian sebelumnya [24] menunjukkan bahwa *SGD* dengan *learning rate* yang lebih besar dapat membantu model keluar dari *sharp minima* dan menuju *flat minima*, yang berkontribusi pada stabilitas generalisasi pada data uji, terutama pada dataset dengan variasi visual yang tinggi seperti citra lesi kulit.

2.4.2. Pelatihan Teacher

Pada proses pemilihan *pretrained teacher* digunakan berbagai arsitektur model yang memiliki jumlah parameter dan ukuran yang besar seperti *DarkNet53*, *ResNet152*, *WideReseNet*, dan *EfficientNet*. Beberapa perbandingan ukuran dan komputasi *pretrained* model pada Tabel 4.

Tabel 4. *Teacher pretrained model.*

No	Pretrained Model	Parameters (M)	GFLOPS
1	*DarkNet-53*	40.59	7.12
2	*ResNet152*	58.16	11.60
3	*RepVitM2-3*	22.41	4.60
4	*EfficientNet-B0*	4.02	0.39
5	*EfficientNet-B3*	10.71	0.97
6	*EfficientNet-B5*	28.35	2.38
7	*WideResNet-50*	66.85	11.45
8	*WideResNet-101*	124.85	22.83

2.4.3. Pelatihan Student

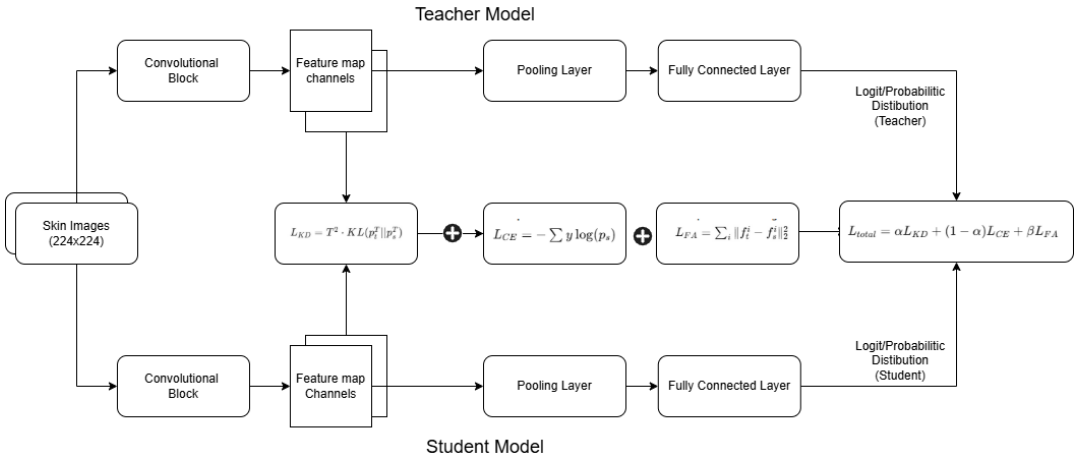
Pada proses pemilihan *pretrained student* digunakan berbagai arsitektur model yang memiliki jumlah parameter dan ukuran yang lebih kecil dan ringan dibandingkan model *teacher* seperti *MobileNetV2*, *ResNet8*, dan *ShuffleNetV2*. Beberapa perbandingan ukuran dan komputasi *pretrained model* pada Tabel 5.

Tabel 5. *Student pretrained model.*

No	Pretrained Model	Parameters (M)	GFLOPS
1	*MobileNetV2*	0.69	0.36
2	*ResNet-8*	0.08	0.63
3	*ShuffleNetV2*	1.26	0.15

2.4.4. Implementasi Hybrid KD

Metode distilasi yang digunakan yaitu *KD Hinton*, *KD Hinton + Hard Label* dan *KD Hinton + Feature Alignment* sekaligus *Hybrid KD*. *Proposed method* ditunjukan pada Gambar 2.



Gambar 2. *Hybrid KD Hinton.*

Pada proses *Hybrid Knowledge Distillation*, gambar kulit berukuran 116x116 terlebih dahulu dimasukkan ke dalam *teacher model* dan *student model* melalui blok konvolusional awal yang mengekstraksi *feature map channels*. Baik *teacher* maupun *student* menghasilkan representasi fitur tingkat menengah yang kemudian digunakan untuk perhitungan *feature alignment loss* (L_{FA}), yaitu selisih kuadrat antara fitur *student*

dan *teacher* pada layer tertentu untuk mendorong *student* meniru representasi *internal teacher*. Setelah itu, kedua model melanjutkan pemrosesan melalui *pooling layer* dan *fully connected layer* untuk menghasilkan distribusi probabilitas (*logit/probability distribution*). Distribusi probabilitas *teacher* digunakan untuk menghitung *soft target loss* berbasis *Kullback–Leibler divergence* (L_{KD}) dengan suhu *temperature* tertentu. Sementara itu, *student* juga dihitung terhadap *ground-truth labels* menggunakan *cross-entropy loss* (L_{CE}). Ketiga komponen *loss* tersebut *soft KD loss* (L_{KD}), *hard label loss* (L_{CE}), serta *feature alignment loss* (L_{FA}) digabungkan menjadi *total loss*, yaitu:

$$L_{\text{total}} = \alpha L_{KD} + (1-\alpha)L_{CE} + \beta L_{FA} \quad (1)$$

Dengan α mengatur kontribusi distilasi logit, $(1-\alpha)$mengatur pembelajaran berbasis label asli, dan β mengendalikan seberapa kuat *student* dipaksa mengikuti representasi fitur *teacher*. Dengan kombinasi ketiga komponen tersebut, *Hybrid KD* mendorong student untuk meniru perilaku *teacher* bukan hanya pada *level output* (logit) tetapi juga pada level label serta struktur fitur internal, sehingga menghasilkan model *student* yang lebih akurat dan stabil. Parameter α dan β berperan penting dalam mengatur keseimbangan antara pembelajaran berbasis *soft targets*, *hard labels*, dan *feature-level supervision*. Parameter α mengontrol kontribusi *logit-based knowledge distillation loss* (L_{KD}) terhadap *cross-entropy loss* (L_{CE}) sedangkan parameter β mengatur kekuatan *feature alignment loss* (L_{FA}) dalam mendorong *student* untuk meniru representasi *internal teacher*. Nilai $\alpha = 0.85$ dipilih untuk memberikan dominasi pada pembelajaran dari *soft targets* teacher, yang mengandung informasi *dark knowledge* berupa hubungan antar kelas yang tidak sepenuhnya tercermin pada label keras (*hard labels*). Sementara itu, parameter β = 0.02 ditetapkan relatif kecil untuk memastikan bahwa *feature alignment loss* berfungsi sebagai *regularizer* tambahan, bukan sebagai tujuan utama optimasi. Berbagai studi [25], [26] menunjukkan bahwa penalti fitur yang terlalu besar dapat menyebabkan *over-constraint*, terutama ketika arsitektur *teacher* dan *student* berbeda, sehingga justru menghambat kemampuan *student* untuk belajar representasi yang optimal.

2.4.5. Evaluasi Model Klasifikasi

Metrik evaluasi merupakan komponen penting untuk menilai kinerja, efisiensi, dan kepraktisan model *deep learning*, khususnya dalam konteks *knowledge distillation* dan perancangan arsitektur ringan. Top-1 *accuracy* mengukur seberapa sering prediksi dengan probabilitas tertinggi yang dihasilkan model sesuai dengan label *ground truth*, yang secara formal dinyatakan sebagai:

$$Top-1\ Acc = \frac{1}{N}\sum_{i=1}^{N} \text{I.}(\arg\max p_i, c = y_i) \quad (2)$$

N adalah jumlah total sampel pengujian, p_i, c adalah probabilitas prediksi model untuk kelas c pada sampel ke -i, y_i adalah label sebenarnya (*ground truth*) dari sampel ke-I, I. adalah fungsi indikator yang bernilai 1 jika kondisi di dalamnya benar dan 0 jika salah.

Selain itu, *Top-5 Accuracy* digunakan untuk mengukur apakah label kebenaran termasuk dalam lima kelas dengan probabilitas tertinggi yang diprediksi oleh model. Metrik ini sangat relevan pada tugas klasifikasi dengan jumlah kelas yang besar.

$$Top-5\ Acc = \frac{1}{N}\sum_{i=1}^{N} \text{I.}(y_i \in Top-5(p_i)) \quad (3)$$

$Top-5(p_i)$ merupakan himpunan lima kelas dengan probabilitas prediksi tertinggi untuk sampel ke-*i*. simbol lainnya memiliki makna yang sama seperti pada perhitungan *Top-1 Accuracy*.

Untuk mengukur kompleksitas model, jumlah parameter dihitung sebagai total keseluruhan bobot yang dapat dipelajari (*learnable weights*) sebagai berikut:

$$\text{Total Parameter Model} = \sum_{l=1}^{L}(K_l\ x\ K_l\ x\ C_l^{in}\ x\ C_l^{out}\ x\ C_l^{out}) \quad (4)$$

L adalah jumlah *layer* pada model, K_l adalah ukuran *kernel*, C_l^{in} merupakan jumlah *channel input*, C_l^{out} adalah jumlah *layer output*. Pada penelitian ini juga melakukan pengukuran kompleksitas komputasi (*G-Flops*) yang merepresentasikan jumlah operasi *floating point* dalam satu kali *forward pass* sebagai berikut:

$$\text{GFLOPs} = \sum_{l=1}^{L}(2\ x\ K_l^2\ x\ C_l^{in}\ x\ C_l^{out}\ x\ H_l\ x\ W_l)\ \div\ 10^9 \quad (5)$$

Information System Journal (INFOS)
Vol. 8, No. 2, November 2025, pp. 215-224
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2025v8i02.2585

K_l adalah ukuran *kernel* konvolusi, C_l^{in} merupakan jumlah *channel input*, C_l^{out} adalah jumlah *layer output*, $H_l \, x \, W_l$ tinggi dan lebar *feature map output*, faktor 2 merepresentasikan operasi perkalian maupun penjumlahan, $\div \ 10^9$ merepresentasikan konversi ke dalam *GFlops*. Sementara itu juga dilakukan pengukuran waktu training dihitung berdasarkan jumlah *epoch*, ukuran *batch*, dan waktu komputasi per-iterasi, yang mencerminkan efisiensi model dalam proses pembelajaran.

$$\text{Total Waktu Training} = \text{E x} \frac{N_{data}}{B} x \ time_{iter} \quad (6)$$

E merupakan jumlah *epochs*, N_{data} merupakan jumlah data yang dilatih, B adalah *batch size*, $time_{iter}$ adalah waktu komputasi per-iterasi (*forward* dan *backward*).

3. HASIL DAN PEMBAHASAN

Hasil eksperimen yang ditampilkan pada Tabel 6., menunjukkan variasi yang signifikan baik dari sisi akurasi maupun biaya komputasi pada berbagai model *teacher* yang dievaluasi. Di antara seluruh arsitektur, *WideResNet-101* mencapai akurasi Top-1 tertinggi sebesar 88.44% dan akurasi Top-5 sebesar 99.08%, yang menegaskan kapasitas representasinya yang sangat kuat. Namun demikian, kinerja tersebut diperoleh dengan waktu komputasi terlama, yaitu 364.20 menit, yang menunjukkan adanya *trade-off* yang jelas antara akurasi dan efisiensi pelatihan.

Tabel 6. Hasil performa *training model teacher.*

Model Teacher	Akurasi (%)		Waktu *Training*
	Top-1	Top-5	(Menit)
DarkNet-53	65.79	97.00	82.82
ResNet152	65.24	96.19	170.56
RepVitM2-3	64.65	95.87	41.90
EfficientNet-B3	81.51	97.88	114.37
EfficientNet-B5	66.62	96.91	222.08
WideResNet-50	87.71	98.54	191.04
WideResNet-101	88.44	99.08	364.20

Sebagai perbandingan, *WideResNet-50* menghasilkan akurasi yang sedikit lebih rendah (Top-1: 87.71%; Top-5: 98.54%), namun dengan waktu komputasi yang jauh lebih singkat (191.04 menit), sehingga memberikan keseimbangan yang lebih baik antara kinerja dan efisiensi. Model yang lebih ringan seperti *RepViTM2-3* dan *DarkNet-53* memiliki waktu komputasi yang jauh lebih cepat (masing-masing 41.90 menit dan 82.82 menit), tetapi akurasi Top-1 yang dicapai relatif lebih rendah, berkisar antara 64.65% – 65.79%, yang mengindikasikan keterbatasan pada kedalaman representasi fitur. *EfficientNet-B3* menonjol sebagai alternatif yang efisien dengan mencapai akurasi Top-1 sebesar 81.51% dan waktu komputasi moderat (114.37 menit), sehingga cocok untuk skenario yang membutuhkan kinerja tinggi dengan keterbatasan sumber daya.

Secara keseluruhan, hasil ini menegaskan bahwa pemilihan model sangat bergantung pada keseimbangan antara akurasi dan biaya komputasi, di mana arsitektur yang lebih dalam cenderung memberikan performa yang lebih baik dengan konsekuensi waktu pelatihan yang lebih lama. Perbandingan kinerja antar model *student* yang ditunjukkan pada Tabel 7., juga memperlihatkan *trade-off* yang serupa.

Tabel 7. Hasil performa *training model student.*

Model Student	Akurasi (%)		Waktu Training
	Top-1	Top-5	(Menit)
MobileNetV2	38.50	97.80	54.42
ResNet-8	65.98	97.17	93.55
ShuffleNetV2	78.86	98.75	213.02
EfficientNet-B0	69.22	97.55	149.86

ShuffleNetV2 mencapai akurasi Top-1 tertinggi sebesar 78.86% dan akurasi Top-5 sebesar 98.75%, yang menunjukkan kemampuannya dalam menyerap representasi dari model *teacher* secara efektif. Namun, pencapaian ini disertai dengan waktu komputasi terlama di antara model *student*, yaitu 213.02 menit, sehingga menjadi model yang paling tidak efisien. Sebaliknya, *MobileNetV2* menunjukkan waktu pelatihan tercepat

(54.42 menit), tetapi memiliki akurasi Top-1 terendah (38.50%), yang mengindikasikan keterbatasan arsitektur ringan dalam menyerap pengetahuan dari *teacher* tanpa proses distilasi. *ResNet-8* dan *EfficientNet-B0* berada di antara dua ekstrem tersebut, di mana *ResNet-8* memberikan akurasi yang lebih baik dibandingkan *MobileNetV2* dengan waktu pelatihan yang *moderate*, sementara *EfficientNet-B0* mencapai akurasi Top-1 yang sedikit lebih tinggi dengan tambahan biaya komputasi. Untuk meningkatkan kinerja model *student* yang ringan, seluruh model tersebut kemudian menjalani proses *knowledge distillation*. Dengan mempertimbangkan keterbatasan sumber daya komputasi, *WideResNet-50* dipilih sebagai model *teacher*, sedangkan *MobileNetV2*, yang memiliki akurasi dan waktu komputasi terendah, dipilih sebagai model student untuk mengevaluasi sejauh mana pengetahuan dapat ditransfer ke arsitektur yang sangat efisien. Beberapa strategi distilasi diterapkan, yaitu *KD Hinton* standar, *KD Hinton* dengan supervisi *Hard Label*, *KD Hinton* dengan tambahan penyelarasan fitur, serta pendekatan *Hybrid KD* yang mengombinasikan seluruh komponen tersebut. Tabel 8 merupakan hasil perbandingan dari *KD Hinton* dan *Hybrid KD* yang diusulkan.

Tabel 8. Perbandingan performa *MobileNetV2: KD Hinton* vs *Hybrid KD Hinton.*

Model Student	*MobileNetV2*			
	Accuracy		*Computation Time*	Parameter (M)
	Top-1	Top-5		
KD Hinton	81.46	99.52	124.16	0.69
KD Hinton + Hard Label	81.98	99.59	124.20	0.69
KD Hinton + Feature Alignment	81.31	99.44	124.25	0.69
Hybrid KD Hinton	82.07	98.80	125.08	0.69

Hasil eksperimen distilasi yang ditampilkan pada Tabel 8., menunjukkan bahwa *KD Hinton* standar telah memberikan peningkatan signifikan pada akurasi Top-1 *MobileNetV2* menjadi 81.46%, yang menegaskan manfaat supervisi *soft label* dari model *teacher*. Penambahan *Hard Label* meningkatkan akurasi Top-1 menjadi 81.98%, yang menunjukkan bahwa informasi *ground truth* berperan penting dalam menstabilkan proses pembelajaran. Sebaliknya, penyelarasan fitur saja sedikit menurunkan akurasi Top-1 menjadi 81.31%, yang mengindikasikan adanya kendala optimasi ketika supervisi ruang fitur diterapkan pada arsitektur yang sangat ringan. Pendekatan *Hybrid KD Hinton*, yang mengombinasikan *Soft Label*, *Hard Label*, dan penyelarasan fitur, menghasilkan kinerja terbaik dengan akurasi Top-1 sebesar 82.07%, tanpa adanya peningkatan jumlah parameter model (tetap sebesar 0.69 juta). Meskipun waktu komputasi meningkat secara marginal, peningkatan akurasi yang diperoleh menunjukkan bahwa strategi distilasi hibrida mampu meningkatkan kualitas representasi model *student* secara efektif tanpa mengorbankan efisiensi, sehingga sangat sesuai untuk aplikasi klasifikasi penyakit kulit berbasis perangkat dengan sumber daya terbatas.

4. KESIMPULAN

Penelitian ini secara sistematis mengevaluasi efektivitas *Hybrid Knowledge Distillation* (Hybrid KD) dalam meningkatkan kinerja model student berarsitektur ringan pada tugas klasifikasi penyakit kulit, dengan mempertimbangkan secara eksplisit akurasi, waktu komputasi, kompleksitas model (parameter), dan biaya komputasi (*GFLOPS*). Eksperimen diawali dengan analisis komparatif berbagai arsitektur *teacher* model berkapasitas besar, mulai dari *EfficientNet* hingga *WideResNet*, untuk mengidentifikasi *trade-off* antara kekuatan representasi dan biaya pelatihan. Hasil pelatihan *teacher* menunjukkan bahwa *WideResNet-101*, dengan 124.85 juta parameter dan 22.83 *GFLOPS*, mencapai performa tertinggi (Top-1 88.44%; Top-5 99.08%), namun memerlukan waktu pelatihan terlama sebesar 364.20 menit, sehingga kurang efisien untuk skenario praktis. Sebaliknya, *WideResNet-50* menghasilkan akurasi yang hampir setara (Top-1 87.71%) dengan waktu pelatihan yang jauh lebih singkat (191.04 menit), sehingga dipilih sebagai *teacher* terbaik dengan keseimbangan optimal antara akurasi dan efisiensi. Model yang lebih ringan seperti *RepViTM2-3* dan *DarkNet-53* memang memiliki waktu pelatihan jauh lebih singkat, namun akurasi Top-1 yang rendah (≈65%) mengindikasikan keterbatasan kapasitas representasi untuk bertindak sebagai *teacher* yang efektif. Pada sisi *student* model, evaluasi awal tanpa distilasi menunjukkan bahwa *MobileNetV2*, meskipun sangat efisien dengan hanya 0.69 juta parameter dan 0.36 *GFLOPS*, memiliki performa Top-1 terendah (38.50%), menegaskan bahwa arsitektur sangat ringan tidak mampu mempelajari representasi diskriminatif secara optimal melalui pembelajaran konvensional. Model student lain seperti *ShuffleNetV2* dan *EfficientNet-B0* mencapai akurasi lebih tinggi, tetapi dengan waktu pelatihan yang meningkat secara signifikan, sehingga mengurangi keunggulan efisiensi. Melalui penerapan *Knowledge Distillation*, khususnya dengan *WideResNet-50* sebagai *teacher* dan *MobileNetV2* sebagai *student*, terjadi peningkatan performa yang sangat signifikan. *KD*

Hinton standar berhasil meningkatkan akurasi Top-1 *MobileNetV2* dari 38.50% menjadi 81.46%, membuktikan efektivitas supervisi *Soft Label* dalam mentransfer pengetahuan. Penambahan *Hard Label supervision* meningkatkan stabilitas pembelajaran dan menghasilkan akurasi Top-1 sebesar 81.98%, sementara *Feature Alignment* secara tunggal tidak memberikan peningkatan optimal, menunjukkan adanya keterbatasan penyelarasan fitur pada arsitektur yang sangat ringan. Pendekatan *Hybrid KD Hinton*, yang mengombinasikan *Soft Label*, *Hard Label*, dan *Feature Alignment*, memberikan hasil terbaik dengan akurasi Top-1 sebesar 82.07%, tanpa adanya peningkatan jumlah parameter model (tetap 0.69 juta parameter). Waktu komputasi hanya meningkat secara marginal (≈125 menit), yang masih jauh lebih efisien dibandingkan pelatihan *teacher* model berkapasitas besar. Hasil ini menegaskan bahwa *Hybrid KD* mampu meningkatkan kualitas representasi internal *student* secara efektif dengan biaya komputasi yang sangat rendah, menjadikannya solusi praktis untuk *deployment* pada perangkat dengan sumber daya terbatas.

Secara keseluruhan, eksperimen ini menunjukkan bahwa *Hybrid Knowledge Distillation* secara efektif meminimalkan *trade-off* antara akurasi dan efisiensi, dengan mentransfer informasi diagnostik penting dari *teacher* berkapasitas besar ke *student* ringan tanpa menambah kompleksitas model saat inferensi. Pendekatan ini sangat relevan untuk aplikasi klasifikasi penyakit kulit berbasis citra pada sistem klinis tertanam dan perangkat *mobile*, di mana keterbatasan memori, waktu inferensi, dan konsumsi daya menjadi faktor krusial. Kedepannya, penelitian ini dapat dikembangkan dengan eksplorasi penyelarasan fitur adaptif, penentuan *layer* distilasi yang lebih selektif, serta evaluasi pada dataset medis yang lebih beragam untuk meningkatkan generalisasi dan keandalan sistem diagnosis berbasis kecerdasan buatan.

DAFTAR PUSTAKA

[1] F. Olayah, E. M. Senan, I. A. Ahmed, and B. Awaji, "AI Techniques of Dermoscopy Image Analysis for the Early Detection of Skin Lesions Based on Combined CNN Features," *Diagnostics*, vol. 13, no. 7, p. 1314, Apr. 2023, doi: 10.3390/diagnostics13071314.
[2] M. Bhalekar and M. Bedekar, "D-CNN: A New model for Generating Image Captions with Text Extraction Using Deep Learning for Visually Challenged Individuals," Engineering, Technology & Applied Science Research, vol. 12, no. 2, pp. 8366–8373, Apr. 2022, doi: 10.48084/etasr.4772.
[3] N. Hameed, A. M. Shabut, and M. A. Hossain, "Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine," in 2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), IEEE, Dec. 2018, pp. 1–7. doi: 10.1109/SKIMA.2018.8631525.
[4] S. Rohhila and A. K. Singh, "Deep learning-based encryption for secure transmission digital images: A survey," Computers and Electrical Engineering, vol. 116, no. February, 2024, doi: 10.1016/j.compeleceng.2024.109236.
[5] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020.
[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
[7] A. Ben Slama, H. Sahli, Y. Amri, and H. Trabelsi, "Res-Net-VGG19: Improved tumor segmentation using MR images based on Res-Net architecture and efficient VGG gliomas grading," Applications in Engineering Science, vol. 16, no. October, 2023, doi: 10.1016/j.apples.2023.100153.
[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 36th International Conference on Machine Learning, ICML 2019, vol. 2019-June, pp. 10691–10700, 2019.
[9] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," NIPS, Mar. 2015.
[10] A. D. Ho, H. G. Doan, and N. T. Nguyen, "Abnormal Human Behavior Detection Improvement with an Efficient Attention Block," Engineering, Technology & Applied Science Research, vol. 15, no. 4, pp. 25048–25054, Aug. 2025, doi: 10.48084/etasr.11463.
[11] P. Bintoro and A. Harjoko, "Lampung Script Recognition Using Convolutional Neural Network," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 16, no. 1, p. 23, Jan. 2022, doi: 10.22146/ijccs.70041.
[12] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks," Multimed. Tools Appl., vol. 79, no. 33–34, pp. 24029–24055, Sep. 2020, doi: 10.1007/s11042-020-09067-2.
[13] C. Flosdorf, J. Engelker, I. Keller, and N. Mohr, "Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer," Aug. 2024.
[14] X. Zhang et al., "DermViT: Diagnosis-Guided Vision Transformer for Robust and Efficient Skin Lesion Classification," Bioengineering, vol. 12, no. 4, p. 421, Apr. 2025, doi: 10.3390/bioengineering12040421.

[15] A. Wang, H. Chen, Z. Lin, J. Han, and G. Ding, "RepViT: Revisiting Mobile CNN From ViT Perspective," in CVPR, Jul. 2023.
[16] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, [Online]. Available: http://arxiv.org/abs/1704.04861
[17] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 6848–6856, 2018, doi: 10.1109/CVPR.2018.00716.
[18] Abdullah s. canipek, "Skin Diseases," https://www.kaggle.com/datasets/ascanipek/skin-diseases/.
[19] Apache MXNet, "Random Horizontal Flip," Papers With Code.
[20] Torch Contributors, "ColorJitter," https://docs.pytorch.org/vision/main/generated/torchvision.transforms.ColorJitter.html.
[21] B. BAKIRARAR and A. H. ELHAN, "Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research," Turkiye Klinikleri Journal of Biostatistics, vol. 15, no. 1, pp. 19–29, 2023, doi: 10.5336/biostatic.2022-93961.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[23] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," Jun. 2017, doi: 10.1109/TMI.2016.2535302.
[24] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima," Feb. 2017.
[25] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for Thin Deep Nets," Dec. 2014.
[26] S. Zagoruyko and N. Komodakis, "Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer," Feb. 2017