Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

Evaluasi Strategi Fine-Tuning pada ConvNeXt dan Swin Transformer untuk Klasifikasi Kanker Kulit

Ahmad Bintang Saputra, Sindhu Rakasiwi*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202214700@mhs.dinus.ac.id, [2,*]sindhu.rakasiwi@dsn.dinus.ac.id
Email Penulis Korespondensi: sindhu.rakasiwi@dsn.dinus.ac.id
Submitted: **22/12/2025**; Accepted: **05/03/2026**; Published**: 06/03/2026**

Abstrak−Kanker kulit merupakan penyakit yang pengidapnya terus meningkat setiap tahun, terutama di wilayah dengan paparan sinar ultraviolet (UV) yang tinggi. Tantangan utama dalam diagnosis kanker kulit terletak pada kemiripan visual antara lesi jinak atau *benign* dan ganas atau *malignant*, sehingga sering menimbulkan kesalahan diagnosis bahkan oleh tenaga medis berpengalaman. Perkembangan teknologi *deep learning* telah memberikan kemajuan signifikan dalam klasifikasi citra medis melalui pendekatan *transfer learning*. Penelitian ini bertujuan untuk membandingkan performa dua arsitektur dari Transformer dan CNN, yaitu Swin Transformer dan ConvNeXt, dalam tugas klasifikasi citra kanker kulit dua kelas *benign* dan *malignant*. Kedua model menggunakan *pretrained* dari ImageNet dan diterapkan dengan tiga strategi *fine-tuning* berbeda, yakni *Linear Probe* (LP), *Full Fine-Tuning* (FT), serta kombinasi dari kedua strategi sebelumnya (LP-FT). Dataset yang digunakan adalah Dataset ISIC Archive dengan pembagian data 80:20 untuk pelatihan dan validasi, dataset terdiri dari 3.297 citra yang terbagi ke dalam dua kelas, dengan 1.800 citra *benign* dan 1.497 citra *malignant*. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa Swin Transformer dengan strategi LP-FT menghasilkan performa terbaik dengan *accuracy* 92,27%, *precision* 92,24%, *recall* 92,17%, dan *F1-score* 92,20%. Temuan ini menunjukkan bahwa pendekatan *fine-tuning* dua tahap mampu meningkatkan stabilitas dan generalisasi model, serta memberikan kontribusi terhadap pengembangan sistem diagnosis kanker kulit berbasis *artificial intelligence* yang lebih akurat.

Kata Kunci: Kanker Kulit; Swin Transformer; ConvNeXt; Fine-Tuning; Linear Probe

Abstract−Skin cancer is one of the diseases whose prevalence continues to increase every year, especially in areas with high exposure to ultraviolet (UV) rays. The main challenge in diagnosing skin cancer lies in the visual similarity between benign and malignant lesions, which often leads to misdiagnosis even by experienced medical personnel. The development of deep learning technology has made significant progress in medical image classification through a transfer learning approach. This study aims to compare the performance of two architectures from Transformer and CNN, namely Swin Transformer and ConvNeXt, in the task of classifying two class benign and malignant skin cancer images. Both models use pretrained from ImageNet and are applied with three different fine-tuning strategies, namely Linear Probe (LP), Full Fine-Tuning (FT), and a combination of the two previous strategies (LP-FT). The dataset used is the ISIC Archive Dataset with an 80:20 data split for training and validation, consisting of 3.297 images divided into two classes, with 1800 benign images and 1.497 malignant images. The evaluation was performed using the accuracy, precision, recall, and F1-score metrics. Swin Transformer with the LP-FT strategy achieved the best performance, with an accuracy of 92,27%, precision of 92,24%, recall of 92,17%, and an F1-score of 92,20%. These findings indicate that the two-stage fine-tuning approach can improve model stability and generalization, as well as contribute to the development of a more accurate artificial intelligence based skin cancer diagnosis system.

Keywords: Skin Cancer; Swin Transformer; ConvNeXt; Fine-Tuning; Linear Probe

1. PENDAHULUAN

Kanker kulit menjadi masalah kesehatan global karena termasuk penyakit yang sering ditemui dan jumlah penderitanya meningkat setiap tahun [1]. Penyebab utamanya paparan sinar ultraviolet (UV) yang berlebihan sehingga memicu kerusakan pada sel kulit [2]. Pada tahun 2021, sekitar 6,64 juta kasus baru kanker kulit dilaporkan secara global berdasarkan estimasi Global Burden of Disease (GBD) [3]. Faktor genetik dan riwayat keluarga juga memainkan peran penting dalam memicu terjadinya kanker kulit, dengan individu yang memiliki keluarga dekat penderita kanker kulit memiliki risiko lebih tinggi [4]. Dalam praktik medis, diagnosis kanker kulit umumnya masih dilakukan melalui pemeriksaan visual oleh dokter kulit, namun akurasinya hanya sekitar 60%. Tantangan utama dalam mendeteksi melanoma adalah kemiripan visual antara lesi jinak atau *benign* dan ganas atau *malignant*, sehingga diagnosis sulit dilakukan bahkan bagi tenaga medis yang berpengalaman [5]. Perkembangan *deep learning* telah membawa kemajuan besar dalam klasifikasi citra medis, terutama melalui pendekatan *transfer learning* pada model *pretrained*. Arsitektur Convolutional Neural Network atau sering disebut CNN seperti ResNet, DenseNet, EfficientNet dan Inception digunakan pada deteksi kanker kulit dan menunjukkan akurasi tinggi pada berbagai dataset dermatoskopik [6]. Namun, CNN umumnya hanya efektif menangkap fitur lokal dan kurang mampu memahami hubungan global antar-piksel yang kompleks pada citra medis [7].

Vision Transformer atau sering disingkat ViT merupakan salah satu model yang memperkenalkan paradigma ini dengan memperlakukan citra sebagai kumpulan *patch* berukuran tetap dan memprosesnya dengan mekanisme *self attention global* seperti pada Natural Language Processing [8]. Sementara itu, Swin Transformer memperkenalkan pendekatan *shifted window attention* yang memungkinkan model mempelajari bagian-bagian citra secara bertahap dan efisien. Pendekatan tersebut membuat Swin mampu memahami informasi lokal dan global sekaligus, tanpa membutuhkan daya komputasi sebesar Vision Transformer. Berbeda dengan mekanisme *global attention* pada Vision Transformer yang menghitung hubungan antar seluruh patch citra dan kurang efisien pada resolusi tinggi, *shifted*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

window attention membatasi perhatian pada jendela lokal untuk menangkap informasi lokal penting seperti tekstur dan batas lesi, serta melakukan pertukaran informasi antar-jendela, sehingga lebih sesuai untuk citra medis beresolusi tinggi seperti kanker kulit. Pendekatan hierarkis yang digunakan juga menjadikan Swin Transformer lebih fleksibel untuk berbagai tugas seperti klasifikasi, deteksi, dan segmentasi citra medis [9]. Sebagai respons terhadap banyaknya arsitektur transformer yang bermunculan, muncul arsitektur ConvNeXt, sebuah upaya untuk memodernisasi CNN agar setara dengan model berbasis Transformer. Pendekatan ini menunjukkan bahwa dengan pembaruan desain dan strategi pelatihan yang tepat, CNN masih dapat bersaing dalam akurasi dan efisiensi tanpa kehilangan sifat konvolusionalnya [10]. Ketiga model tersebut menggambarkan evolusi dari arsitektur CNN dan Transformer.

Sejalan dengan perkembangan arsitektur modern, sejumlah studi telah melaporkan implementasi langsung dari model-model tersebut pada citra medis, Pendekatan *deep transfer learning* berbasis CNN juga terbukti efektif untuk klasifikasi kanker kulit. Model *pretrained* ResNet101 dan DenseNet201 yang diuji pada dataset ISIC 2020 menunjukkan performa tinggi, di mana ResNet101 mencapai akurasi 87% dan AUC 0.943, lebih unggul dibanding DenseNet201 [11]. Newaz et al. [12] membandingkan berbagai arsitektur CNN dan Transformer dengan menggunakan data multikelas penyakit kulit berbasis citra yang diambil melalui ponsel, hasil penelitian menunjukkan bahwa Swin Transformer memberikan hasil terbaik dengan akurasi 86%, mengungguli model CNN seperti ResNet dan Inception. Dagnaw et al. [13] melaporkan bahwa model Vision Transformer (ViT) dan Swin Transformer memberikan performa tinggi untuk klasifikasi kanker kulit, dengan ViT mencapai akurasi 88,6% dan Swin tiny 87,7%. Hasil ini menunjukkan efektivitas arsitektur transformer dalam mengenali pola lesi kanker kulit secara akurat dan efisien. Sementara itu, Ahmed et al. [14] menilai performa beberapa model Transformer termasuk Swin Transformer untuk klasifikasi multikelas penyakit kulit menggunakan berbagai kombinasi dataset. Studi tersebut menegaskan bahwa Swin Transformer menunjukkan performa unggul dibandingkan model CNN seperti DenseNet201 dan EfficientNetB5, terutama pada citra dengan kesamaan visual yang sangat mirip antar kelas. Selain itu, implementasi model ConvNeXt untuk klasifikasi penyakit kulit, hasil eksperimen pada dataset HAM10000 menunjukkan bahwa ConvNeXt mencapai akurasi 87.62% [15].

Dalam penerapan *transfer learning*, terdapat beberapa strategi *fine-tuning* yang umum digunakan, yaitu *Linear Probe* (LP) dan *Full Fine-Tuning* (FT). LP hanya melatih lapisan klasifikasi terakhir dengan membekukan seluruh *backbone* model *pretrained*, sehingga lebih efisien dan stabil pada dataset berukuran terbatas. Sebaliknya, FT melatih seluruh lapisan model untuk memungkinkan adaptasi fitur yang lebih menyeluruh terhadap domain citra medis, meskipun dengan risiko *overfitting* yang lebih tinggi. Untuk mengombinasikan keunggulan keduanya, digunakan strategi LP-FT, yaitu pendekatan dua tahap yang diawali dengan LP, kemudian dilanjutkan dengan FT. Strategi ini memungkinkan adaptasi model yang lebih optimal dan stabil dalam mengenali pola kompleks pada citra kanker kulit [16], [17].

Penelitian ini bertujuan untuk mengevaluasi pengaruh strategi *fine-tuning* dalam penerapan *transfer learning* pada klasifikasi citra kanker kulit, dengan menggunakan ConvNeXt dan Swin Transformer sebagai representasi dua paradigma utama dalam *computer vision* modern, yaitu CNN modern dan Transformer hierarkis. Swin Transformer merepresentasikan arsitektur berbasis *self-attention* yang mampu menangkap hubungan global dan lokal secara hierarkis, sedangkan ConvNeXt merepresentasikan CNN modern yang mengadopsi prinsip desain Transformer. Kedua model dipilih dalam varian yang setara serta menggunakan *pretrained weights* dari ImageNet dengan tingkat kompleksitas dan kapasitas representasi yang sebanding, sehingga perbandingan yang dilakukan bersifat adil dan seimbang.

Meskipun berbagai penelitian terdahulu telah melaporkan kinerja tinggi model CNN dan Transformer pada klasifikasi kanker kulit, sebagian besar studi masih menerapkan pendekatan *Full fine-tuning* (FT) dan berfokus pada hasil akhir tanpa mengevaluasi secara sistematis pengaruh strategi *fine-tuning* terhadap stabilitas pelatihan dan kemampuan generalisasi model, khususnya pada dataset medis. *Full fine-tuning* (FT) berpotensi menyebabkan *overfitting* atau *negative transfer* pada kondisi tersebut [17], sedangkan pendekatan bertahap seperti LP-FT dilaporkan mampu meningkatkan stabilitas pelatihan melalui adaptasi model secara bertahap [16]. Hal ini menjadi penting pada klasifikasi citra kanker kulit yang memiliki kompleksitas visual tinggi, kemiripan antar-kelas, serta perbedaan domain dengan citra alami non-medis yang digunakan pada model *pretrained* seperti ImageNet.

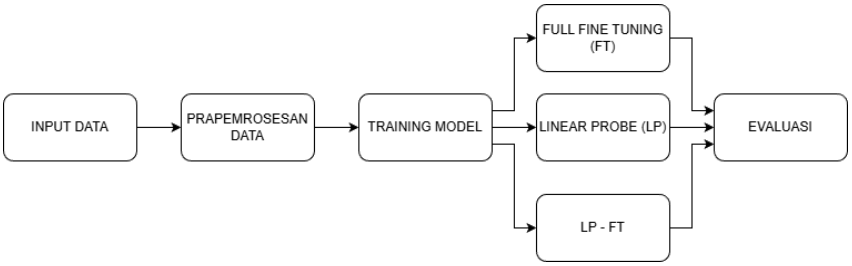
Berdasarkan hal tersebut, penelitian ini mengisi celah penelitian dengan melakukan evaluasi terhadap strategi LP, FT, dan LP-FT pada ConvNeXt dan Swin Transformer dalam klasifikasi citra kanker kulit. Penelitian ini memberikan kontribusi berupa analisis komparatif antara Swin Transformer dan ConvNeXt pada citra medis kanker kulit serta penentuan strategi *fine-tuning* yang efektif untuk meningkatkan stabilitas pelatihan dan kinerja model, sehingga dapat menjadi acuan dalam pengembangan sistem diagnosis kanker kulit berbasis *artificial intelligence*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan *transfer learning* dengan memanfaatkan model Swin Transformer dan ConvNeXt yang telah dilatih sebelumnya pada dataset ImageNet-1K atau yang biasa disebut *pretrained* model. Kemudian disesuaikan melalui proses *fine-tuning* untuk melakukan klasifikasi dua kelas citra kanker kulit, yaitu *benign* dan *malignant*. Tahapan penelitian ini dapat dilihat pada Gambar 1.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

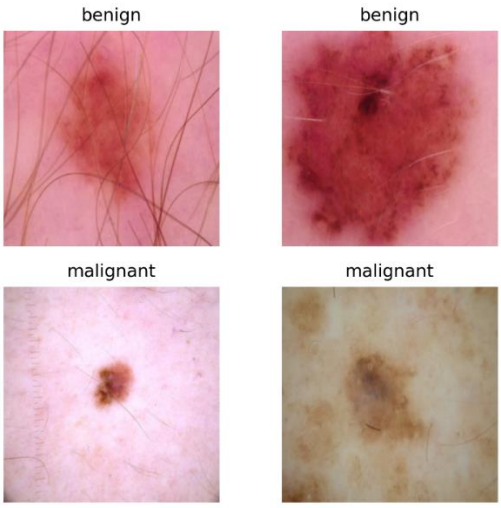


Gambar 1. Alur tahapan penelitian

Gambar 1 menunjukkan alur tahapan penelitian yang dilakukan secara sistematis mulai dari tahap input data hingga evaluasi model. Proses diawali dengan pengumpulan data citra kanker kulit yang selanjutnya melalui tahap prapemrosesan untuk meningkatkan kualitas dan keragaman data, meliputi *resizing*, augmentasi, serta normalisasi citra. Data yang telah diprapemroses kemudian digunakan pada tahap pelatihan model dengan pendekatan *transfer learning* menggunakan model *pretrained*. Pada tahap ini, pelatihan dilakukan dengan tiga strategi berbeda, yaitu *Linear Probe* (LP), *Full Fine-Tuning* (FT), dan kombinasi keduanya (LP-FT), untuk menganalisis pengaruh masing-masing strategi terhadap performa model. Seluruh model yang dihasilkan dari ketiga strategi tersebut selanjutnya dievaluasi menggunakan metrik *accuracy*, *precision*, *recall, dan F1-score* dan confusion matrix guna menilai kemampuan model dalam mengklasifikasikan citra kanker kulit.

2.2 Pengumpulan Data

Penelitian ini menggunakan dataset publik Kaggle yang diperoleh dari ISIC Archive. Dataset terdiri dari 3.297 citra yang terbagi ke dalam dua kelas, dengan 1.800 citra *benign* dan 1.497 citra *malignant*. Gambar 2 menunjukkan sampel citra dari kedua kelas, yaitu *benign* dan *malignant.*



Gambar 2. Sampel citra

Gambar 2 menampilkan contoh citra kanker kulit dari dua kategori kelas pada dataset. Secara visual terlihat adanya variasi bentuk, ukuran area, pola warna, tekstur, serta tingkat kontras antar citra. Citra kanker kulit jinak (*benign*) biasanya menunjukkan pola visual yang lebih seragam, dengan warna yang relatif konsisten, bentuk yang cenderung simetris, serta batas area yang lebih halus dan terdefinisi dengan baik. Distribusi pigmen pada citra jinak umumnya terlihat lebih merata dan tidak memperlihatkan kontras yang ekstrem. Sebaliknya, citra kanker kulit ganas (*malignant*) cenderung menampilkan variasi warna yang lebih beragam dalam satu area, bentuk yang tidak simetris, tepi yang tidak beraturan, serta pola pigmen yang tersebar tidak merata. Perbedaan karakteristik visual tersebut menjadi dasar bagi model klasifikasi citra dalam mempelajari pola pembeda antar kelas.

2.3 Prapemrosesan Data

Pada tahap ini dilakukan serangkaian transformasi data untuk meningkatkan keragaman citra dan membantu model dalam melakukan generalisasi pada saat pelatihan model.

Tabel 1. Pembagian dataset

Kelas	Data Latih	Data Validasi	Total
Benign	1400	360	1800
Malignant	1197	300	1497

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

Tabel 1 berisi pembagian dataset yang digunakan, data dibagi menjadi dua bagian yaitu data pelatihan dan data validasi, dibagi dengan perbandingan 80:20. 80% dari total citra digunakan untuk pelatihan dan 20% untuk validasi. Proses pembagian data dilakukan menggunakan *stratified split*, sehingga proporsi setiap kelas pada data pelatihan dan validasi tetap seimbang dan merepresentasikan distribusi kelas pada dataset asli.

Teknik augmentasi data diterapkan hanya pada data pelatihan untuk meningkatkan keragaman citra, proses ini mencakup *resizing* citra menjadi ukuran 224 × 224 piksel, kemudian penerapan augmentasi seperti *random horizontal flip*, *random vertical flip*, serta rotasi acak hingga 15 derajat, agar model lebih banyak belajar terhadap variasi citra kulit. Selain itu, penyesuaian warna dengan perubahan kecil pada tingkat kecerahan, kontras dan saturasi masing-masing sebesar 0,15, serta penyesuaian hue sebesar 0,05. Setiap citra baik pada data pelatihan maupun data validasi kemudian dikonversi menjadi tensor dan dinormalisasi nilai pikselnya yang sesuai dengan normalisasi ImageNet *pretrained* model. Proses augmentasi diterapkan pada setiap *batch* selama tahap pelatihan model, sehingga variasi citra yang digunakan dapat berbeda pada setiap iterasi pelatihan. Pendekatan ini dilakukan untuk mengurangi risiko *overfitting* akibat keterbatasan jumlah data.

2.4 Model Deep Learning

Penelitian ini menerapkan pendekatan *transfer learning* untuk memanfaatkan model *pretrained* yang telah dilatih pada dataset berskala besar seperti ImageNet-1K. Pendekatan ini dilakukan untuk mengatasi keterbatasan jumlah data pelatihan yang umum dijumpai pada domain medis serta mempercepat proses konvergensi model [18], [19]. Model yang digunakan meliputi Swin Transformer varian tiny dan ConvNeXt varian tiny. Varian tiny dipilih karena lebih efisien secara komputasi dan sesuai untuk dataset medis berukuran terbatas, kedua varian ini juga memiliki jumlah parameter yang sebanding, masing-masing sekitar 28 juta parameter, sehingga perbandingan dapat dilakukan secara lebih adil. Masing-masing model mewakili arsitektur Transformer hierarkis, dan CNN modern bergaya Transformer. Meskipun penelitian terdahulu menunjukkan potensi yang baik dari kedua arsitektur tersebut dalam klasifikasi citra medis, evaluasi komparatif tetap diperlukan untuk memahami perbedaan kinerja serta menentukan strategi *fine-tuning* yang paling efektif dalam memaksimalkan performa model pada klasifikasi kanker kulit.

2.4.1 Swin Transformer

Swin Transformer dipilih karena kemampuannya dalam menangkap hubungan spasial lokal dan global secara efisien melalui mekanisme *shifted window self attention*. Arsitektur ini mengorganisasi representasi citra secara hierarkis, mirip dengan CNN, namun tetap mempertahankan *global context modeling* dari Transformer [9].

2.4.2 ConvNeXt

ConvNeXt dipilih karena mewakili generasi terbaru CNN yang mengadopsi prinsip desain dari Transformer, seperti *Layer Normalization* dan *GELU activation*. Model ini mempertahankan efisiensi konvolusional dan memperoleh kemampuan generalisasi global yang dimiliki oleh Transformer [10].

2.5 Strategi Fine-tuning dan Hyperparameter

Model dalam penelitian ini dilatih menggunakan tiga strategi pelatihan yang berbeda, yaitu *Linear Probe* (LP), *Full Fine-Tuning* (FT), dan kombinasi strategi *Linear Probe* diikuti *Full Fine-Tuning* (LP-FT). Penggunaan ketiga strategi itu untuk mengadaptasi model *pretrained* ke dataset citra medis.

2.5.1 *Full Fine-Tuning* (FT)

Strategi *Full Fine-Tuning* (FT) melibatkan pelatihan seluruh *layer* model baik *head* maupun *backbone* sejak awal. Cara ini memungkinkan model menyesuaikan diri sepenuhnya terhadap domain baru seperti citra medis, namun dapat menyebabkan penurunan performa atau *overfitting* [16].

2.5.2 *Linear Probe* (LP)

Pada strategi Linear Probe (LP), hanya bagian *classification head* atau lapisan terakhir model yang dilatih sementara *backbone* dibekukan. Pendekatan ini memanfaatkan representasi fitur umum dari model *pretrained* untuk menyesuaikan model terhadap tugas baru tanpa perlu mengubah struktur representasi awal [17].

2.5.3 (LP-FT)

Strategi LP-FT menggabungkan kedua strategi LP dan FT, model terlebih dahulu dilatih dengan LP untuk memperoleh stabilitas dan pemetaan awal yang baik, kemudian dilanjutkan dengan FT untuk menyesuaikan seluruh representasi terhadap karakteristik domain target. Pendekatan dua tahap ini terbukti lebih stabil dan menghasilkan performa lebih baik pada dataset citra medis [16].

2.5.4 Hyperparameter

Pelatihan model dilakukan dengan menggunakan *optimizer* AdamW yang dilengkapi mekanisme *weight decay* untuk mengurangi risiko *overfitting*. *Scheduler learning rate* menggunakan CosineAnnealing scheduler agar konvergensi pelatihan lebih stabil. *Loss function* yang digunakan adalah CrossEntropyLoss dengan penerapan *label smoothing*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

untuk meningkatkan kemampuan generalisasi model. Selain itu, dropout juga diterapkan guna mencegah *overfitting* selama proses pelatihan. Hyperparameter yang digunakan pada semua model tertera pada Tabel 2.

Tabel 2. Hyperparameter yang digunakan

Hyperparameter	Nilai		
	LP	FT	LP-FT
Batch size	32	32	32
Epoch	30	30	Stage 1: 10 Stage 2: 20
Optimizer	AdamW	AdamW	AdamW
Learning rate head	1e-4	1e-4	1e-4
Learning rate backbone	2e-6	2e-6	2e-6
Weight decay	1e-4	1e-4	1e-4
Scheduler	CosineAnnealingLR	CosineAnnealingLR	CosineAnnealingLR
CrossEntropyLoss (label smoothing)	0.05	0.05	0.05
Dropout (head)	0.5	0.5	0.5

Pelatihan untuk strategi LP-FT dilakukan dalam dua tahap, yaitu tahap pertama selama 10 *epoch* dan tahap kedua 20 *epoch* sementara strategi LP dan FT hanya 1 tahap dengan 30 *epoch,* dengan *learning rate* 1e-4 pada bagian *head*, serta 2e-6 untuk *backbone*. *Optimizer* yang digunakan adalah AdamW dengan *weight decay* sebesar 1e-4 untuk mencegah *overfitting*. Penjadwalan *learning rate* diatur menggunakan CosineAnnealing agar proses konvergensi lebih stabil. *Loss function* yang digunakan adalah *CrossEntropyLoss* dengan *label smoothing* sebesar 0.05 untuk meningkatkan generalisasi model, sedangkan *dropout* sebesar 0.5 diterapkan pada bagian *head* untuk mengurangi risiko *overfitting*.

2.6 Evaluasi

Model yang telah dilatih kinerjanya dievaluasi menggunakan confusion matrix dan metrik evaluasi seperti accuracy, precision, recall, dan F1-score. Confusion matrix memberikan gambaran yang jelas seberapa baik model dapat membedakan kelas, dengan membandingkan label yang diprediksi dan label yang sebenarnya, menunjukkan jumlah prediksi yang benar *True Positive* (TP), *True Negative* (TN) dan prediksi yang salah *False Positive* (FP), *False Negative* (FN). *Accuracy* mengukur seberapa akurat model dapat mengklasifikasi data (1), *precision* menunjukkan perbandingan dari sampel positif yang benar dari keseluruhan prediksi positif (2), *recall* menunjukkan model mengenali semua sampel positif yang sebenarnya (3), dan *F1-score* menggabungkan *precision* dan *recall* untuk memperoleh nilai keseimbangan (4). Keempat metrik ini saling melengkapi dalam menilai kinerja model [20][21]. Dalam konteks medis, khususnya klasifikasi kanker kulit, *recall* memiliki peran yang sangat penting karena berhubungan langsung dengan ketepatan diagnosis dan keputusan penanganan pasien, nilai *recall* yang tinggi membantu meminimalkan risiko kanker ganas yang tidak teridentifikasi [22].

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan

Model yang sudah dilatih lalu dievaluasi performanya menggunakan metrik *accuracy*, *precision*, *recall* dan *F1-score*. Evaluasi performa model klasifikasi umumnya menggunakan empat metrik utama, yaitu *accuracy*, *precision*, *recall* dan *F1-score*Hasil pelatihan disajikan pada Tabel 3. Dilanjutkan dengan mengevaluasi kurva *accuracy* dan *loss* lalu menganalisis confusion matrix.

Tabel 3. Hasil Pelatihan

Model	Accuracy	Precision	Recall	F1-score
Swin LP	0.8318	0.8296	0.8320	0.8305
ConvNeXt LP	0.8530	0.8509	0.8509	0.8518
Swin FT	0.9076	0.9058	0.9083	0.9068
ConvNeXt FT	0.9030	0.9014	0.9029	0.9021

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240–2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

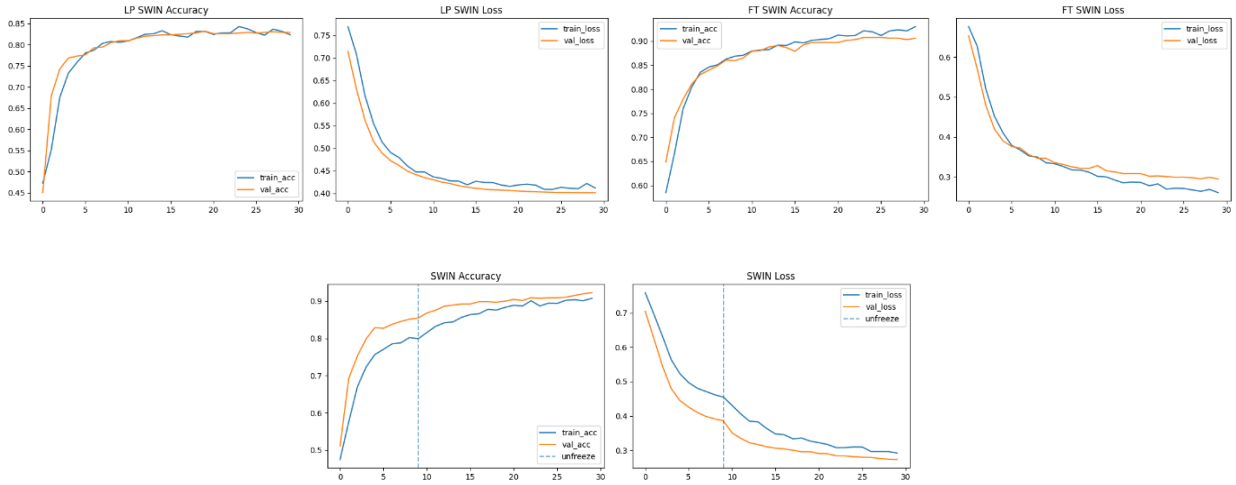
Model	Accuracy	Precision	Recall	F1-score
Swin LP-FT	0.9227	0.9224	0.9217	0.9220
ConvNeXt LP-FT	0.9045	0.9045	0.9028	0.9036

Strategi LP menunjukkan performa yang paling rendah dibandingkan dua strategi lainnya. Model Swin Transformer LP hanya mencapai akurasi 0.8318 dengan *recall* 0.8320, sedangkan ConvNeXt LP sedikit lebih baik dengan akurasi 0.8530 dan *recall* 0.8509. Strategi FT menghasilkan peningkatan yang cukup baik pada kedua model. FT mencapai akurasi 0.9076 dan *recall* 0.9083, sedikit mengungguli ConvNeXt FT dengan akurasi 0.9030 dan *recall* 0.9029.

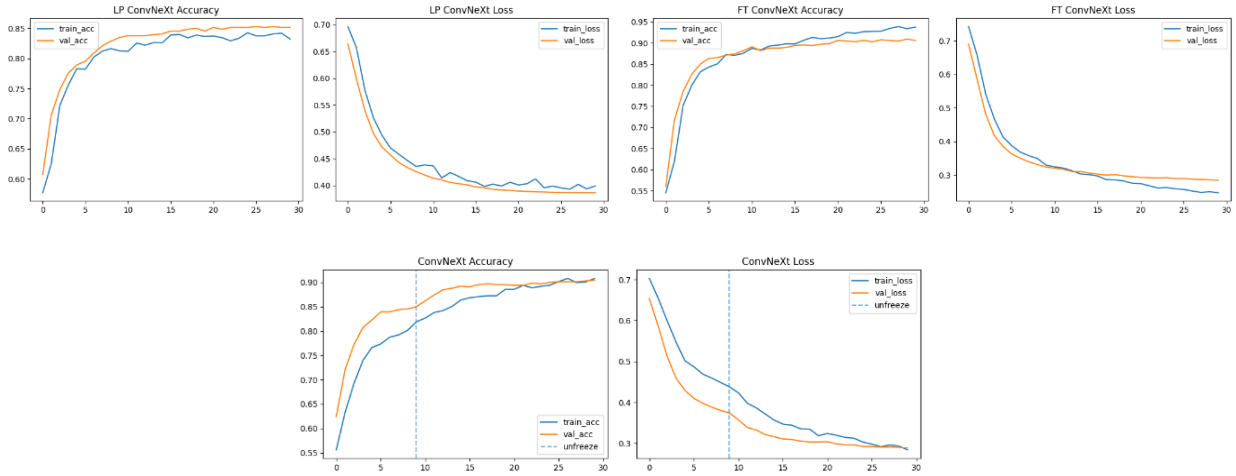
Model Swin dengan strategi LP-FT menunjukkan performa tertinggi dari model lainnya mendapatkan akurasi 0.9227 dan *recall* 0.9217, diikuti oleh ConvNeXt LP-FT dengan akurasi 0.9045 dan *recall* 0.9028. Strategi ini terbukti paling efektif karena mampu memanfaatkan stabilitas representasi dari tahap LP sekaligus menyempurnakannya melalui penyesuaian penuh pada tahap FT, sehingga menghasilkan generalisasi yang lebih baik terhadap data kanker kulit.

3.2 Analisis Grafik Pelatihan

Pada Gambar 3 dan Gambar 4 ditampilkan kurva akurasi dan *loss* dari dua model dan tiga strategi dengan 30 *epoch*, kurva tersebut digunakan sebagai indikator dalam mengevaluasi model yang dilatih menunjukkan potensi *overfitting* atau *underfitting* [23].



Gambar 3. Grafik pelatihan model Swin



Gambar 4. Grafik pelatihan model ConvNeXt

3.2.1 Analisis Strategi *Liner Probe* (LP)

Berdasarkan hasil visualisasi kurva akurasi dan *loss* pada gambar di atas, seluruh model menunjukkan tren pelatihan yang stabil dan konvergen. Pada strategi LP, baik Swin Transformer maupun ConvNeXt memperlihatkan peningkatan akurasi yang konsisten hingga mendekati titik konvergensi, dengan *validation accuracy* yang relatif sejajar dengan *training accuracy*, menandakan bahwa model tidak mengalami *overfitting* yang berarti. Kurva *loss* juga menurun dengan halus tanpa fluktuasi besar, menunjukkan proses optimasi yang stabil. Hal ini mengindikasikan bahwa

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

meskipun hanya *classification head* yang dilatih, representasi fitur dari *pretrained* cukup baik untuk melakukan klasifikasi. Namun, performa LP secara keseluruhan masih terbatas karena backbone tidak disesuaikan dengan karakteristik domain baru, sehingga akurasinya cenderung rendah dari strategi lain. Hal ini menunjukkan bahwa meskipun hanya bagian *head* atau lapisan yang dilatih, model tetap mampu mendapatkan hasil yang baik.

3.2.2 Analisis Strategi *Full Fine-Tuning* (FT)

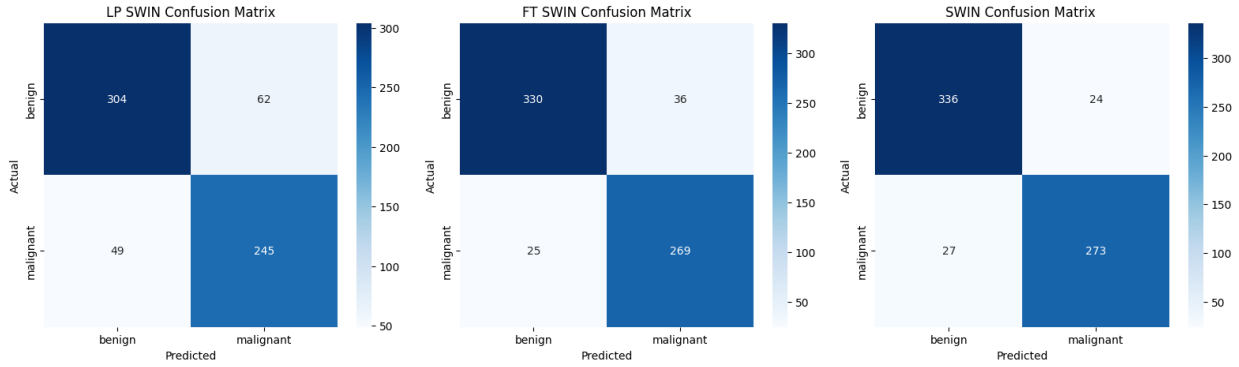
Pada strategi FT, performa kedua model meningkat secara signifikan dibandingkan LP. Baik Swin Transformer FT maupun ConvNeXt FT menunjukkan kurva yang stabil hingga akhir pelatihan. Namun, pada akhir *epoch* tampak sedikit gap antara kurva train dan validasi, terutama pada Swin Transformer, yang mengindikasikan adanya gejala *overfitting* ringan, di mana model belajar sangat baik pada data latih namun sedikit kehilangan generalisasi pada data validasi. Meski demikian, akurasi tetap tinggi dan stabil, menunjukkan bahwa model masih memiliki kemampuan generalisasi yang baik.

3.2.3 Analisis Strategi LP-FT

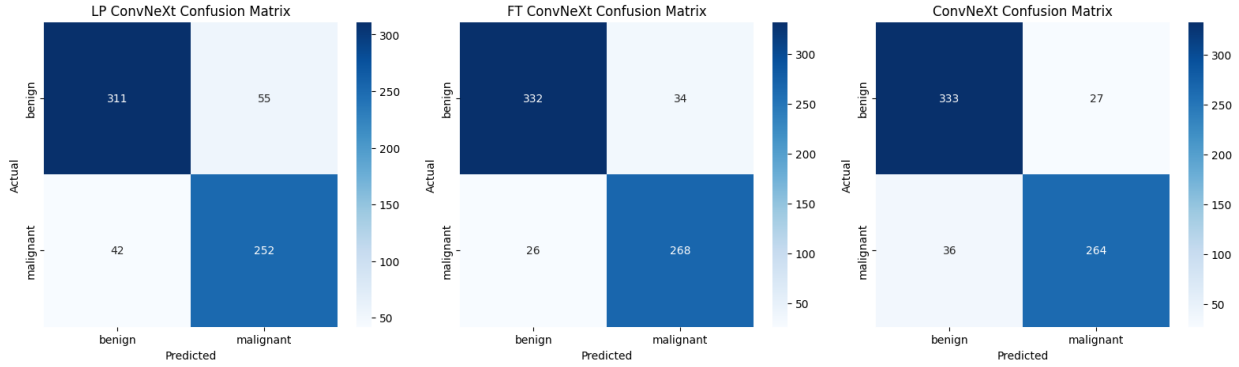
Pada strategi dua tahap LP-FT, yang ditandai dengan garis vertikal pada epoch ke-10 sebagai titik *unfreeze*, terjadi peningkatan performa yang cukup signifikan setelah backbone diaktifkan kembali untuk dilatih. Kedua model menunjukkan lonjakan akurasi serta penurunan *loss* yang konsisten hingga akhir pelatihan, dan tidak menunjukkan tanda-tanda *overfitting*. Kurva tersebut menegaskan bahwa pendekatan LP-FT mampu menyeimbangkan stabilitas representasi dari tahap awal dengan kemampuan adaptasi dari tahap *fine-tuning*, menghasilkan kinerja paling optimal pada tugas klasifikasi citra kanker kulit

3.3 Analisis Confusion Matrix

Confusion matrix memberikan gambaran yang jelas seberapa baik model dapat membedakan kelas, dengan membandingkan label yang diprediksi dan label yang sebenarnya, menunjukkan jumlah prediksi yang benar *true positive* (TP), *true negative* (TN) dan prediksi yang salah *false positive* (FP), *false negative* (FN). Gambar 5 dan Gambar 6 disajikan confusion matrix dari dua model dan tiga strategi.



Gambar 5. Confusion matrix model Swin



Gambar 6. Confusion matrix model ConvNeXt

3.3.1 Analisis Confusion Matrix *Liner Probe* (LP)

Pada strategi Linear Probe, baik Swin Transformer maupun ConvNeXt menunjukkan performa klasifikasi yang baik namun belum optimal karena hanya *classification head* yang dilatih, sementara backbone tetap dibekukan. Untuk Swin Transformer LP, model menghasilkan 304 prediksi benar untuk kelas *benign* dan 245 prediksi benar untuk kelas *malignant*. Sementara itu, terdapat 62 *False Positive (FP)* dan 49 *False Negative (FN)*. Pola ini menunjukkan bahwa

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

meskipun model sudah mampu mengenali sebagian besar citra dengan benar, masih terdapat kecenderungan untuk salah mengenali lesi *malignant* sebagai *benign*. Sedangkan ConvNeXt LP menunjukkan hasil yang sedikit lebih baik dibanding Swin LP dengan 311 *True Positive* untuk *benign* dan 252 *True Positive* untuk *malignant*. Jumlah kesalahan yang terjadi adalah 55 *False Positive* dan 42 *False Negative*.

3.3.2 Analisis Confusion Matrix *Full Fine-Tuning* (FT)

Model Swin Transformer FT mencatat 330 *True Positive* untuk kelas *benign* dan 269 untuk *malignant*. Kesalahan prediksi sangat kecil dengan hanya 36 *False Positive* dan 25 *False Negative*. Distribusi ini menunjukkan keseimbangan yang baik antara kemampuan mengenali lesi jinak dan ganas, dengan tingkat generalisasi yang tinggi. Pola ini juga mencerminkan stabilitas pelatihan pada tahap FT. Sementara ConvNeXt FT juga menunjukkan performa yang kompetitif dengan 332 *True Positive* untuk *benign* dan 268 untuk *malignant*, serta hanya 34 *False Positive* dan 26 *False Negative*. Meskipun hasilnya sedikit di bawah Swin transformer FT, model ini tetap menunjukkan konvergensi yang baik dan kesalahan prediksi yang rendah.

3.3.3 Analisis Confusion Matrix LP-FT

Pada strategi LP-FT (*Linear Probe – Fine-Tuning*), Swin Transformer LP-FT berhasil mengklasifikasikan 336 citra *benign* dan 273 citra *malignant* secara benar, dengan hanya 24 kesalahan *False Positive* dan 27 *False Negative*. Sementara itu, ConvNeXt LP-FT mencatat hasil yang sebanding, dengan 333 prediksi benar pada kelas *benign* dan 264 pada kelas *malignant*, serta 27 *False Positive* dan 36 *False Negative*. Hasil ini memperlihatkan bahwa strategi dua tahap secara efektif meminimalkan kesalahan prediksi pada kedua kelas dan memperkuat generalisasi model terhadap variasi citra lesi kulit. Temuan ini juga menunjukkan bahwa Swin Transformer LP-FT menghasilkan jumlah *False Negative* terendah, yang sangat krusial dalam konteks diagnosis kanker kulit karena kesalahan ini berpotensi menyebabkan kanker ganas terdiagnosis sebagai jinak. Secara keseluruhan, pendekatan LP-FT terbukti meningkatkan stabilitas dan kemampuan adaptasi model *pretrained* pada domain medis, menghasilkan performa klasifikasi yang lebih konsisten dan akurat dibandingkan strategi LP atau FT.

3.4 Pembahasan

Berdasarkan hasil eksperimen yang telah dipaparkan pada bagian sebelumnya, Strategi LP-FT menunjukkan performa terbaik karena menggabungkan stabilitas representasi fitur pada tahap *Linear Probe* dengan kemampuan adaptasi menyeluruh pada tahap *Fine-Tuning*. Pada tahap awal, *backbone* dibekukan sehingga model memanfaatkan representasi fitur dari *pretrained* cukup baik untuk melakukan klasifikasi. Pendekatan ini menjaga struktur fitur dasar tetap stabil dan mengurangi risiko distorsi representasi awal akibat keterbatasan data medis. Selanjutnya, pada tahap *Full Fine-Tuning*, parameter *backbone* dibuka sehingga model dapat beradaptasi lebih spesifik terhadap karakteristik citra dermatoskopik. Temuan ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa *Fine-Tuning* dua tahap dapat menghasilkan performa yang lebih baik dibandingkan strategi lainnya [16]. Hal ini menjelaskan mengapa Swin Transformer dengan strategi LP-FT menghasilkan kinerja paling unggul pada seluruh metrik evaluasi.

Di sisi lain, strategi *Full Fine-Tuning* menunjukkan karakteristik pelatihan yang berbeda. Indikasi *overfitting* ringan pada strategi *Full Fine-Tuning* muncul karena seluruh parameter model diperbarui sejak awal pelatihan, sementara ukuran dataset yang digunakan relatif terbatas, yaitu 3.297 citra. Model modern seperti Swin Transformer dan ConvNeXt memiliki kapasitas parameter yang besar walaupun menggunakan varian yang paling kecil, sehingga sangat fleksibel dalam menyesuaikan diri terhadap data latih. Ketika seluruh layer langsung dibuka, model cenderung terlalu menyesuaikan diri terhadap data pelatihan, sehingga peningkatan akurasi pelatihan tidak selalu diikuti peningkatan yang setara pada data validasi. Kondisi ini diperkuat oleh perbedaan domain antara data *pretrained* ImageNet dan citra dermatoskopik, di mana *Full Fine-Tuning* yang terlalu agresif dapat menggeser representasi fitur umum yang masih relevan, ini sejalan dengan temuan penelitian sebelumnya yang menyatakan bahwa *fine-tuning* penuh dapat menyebabkan distorsi fitur *pretrained* dan menurunkan kemampuan generalisasi pada kondisi data terbatas [17].

Jika dibandingkan dengan penelitian terdahulu, hasil yang diperoleh dalam penelitian ini menunjukkan peningkatan performa yang signifikan. Penelitian sebelumnya melaporkan akurasi Swin Transformer pada klasifikasi kanker kulit berada pada kisaran 86% hingga 87%, sama dengan ConvNeXt yang mencapai akurasi 87%. Pencapaian akurasi 92,27% dalam penelitian ini menunjukkan bahwa strategi *fine-tuning* memiliki peran yang penting dalam memaksimalkan potensi arsitektur modern.

Secara keseluruhan, analisis kurva akurasi, *loss*, dan confusion matrix menunjukkan bahwa seluruh model mengalami proses pelatihan yang konvergen dan stabil. Perbedaan utama terlihat pada pola generalisasi antar strategi *fine-tuning*. Strategi *Linear Probe* (LP) memberikan kestabilan pelatihan yang baik namun peningkatan performanya terbatas, Strategi LP-FT memperlihatkan hasil paling baik antara peningkatan akurasi dan kestabilan kurva validasi, sedangkan strategi *Full Fine-Tuning* menunjukkan gap kecil antara kurva pelatihan dan validasi pada *epoch* akhir sebagai indikasi *overfitting* ringan. Meskipun demikian, tingkat kesalahan klasifikasi pada confusion matrix tetap rendah dan distribusi prediksi antar kelas cukup seimbang. Temuan ini menunjukkan bahwa strategi LP-FT mampu

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

menjaga stabilitas pelatihan sekaligus meningkatkan performa klasifikasi. Dengan demikian, strategi LP-FT memberikan hasil terbaik diantara kedua model dan kemampuan generalisasi pada klasifikasi citra kanker kulit.

4. KESIMPULAN

Penelitian ini berfokus pada evaluasi strategi *fine-tuning* dalam penerapan *transfer learning* untuk klasifikasi citra kanker kulit dua kelas *benign* dan *malignant*, dengan menggunakan arsitektur ConvNeXt dan Swin Transformer sebagai representasi CNN modern dan Transformer hierarkis. Kedua arsitektur dievaluasi menggunakan tiga strategi pelatihan, yaitu *Linear Probe* (LP), *Full Fine-Tuning* (FT), dan pendekatan dua tahap *Linear Probe–Fine-Tuning* (LP-FT), untuk menganalisis pengaruh strategi *fine-tuning* terhadap performa, stabilitas pelatihan, dan kemampuan generalisasi model. Hasil eksperimen menunjukkan bahwa Swin Transformer dengan strategi LP-FT unggul dengan menghasilkan performa terbaik dengan *accuracy* 92,27%, *precision* 92,24%, *recall* 92,17%, dan *F1-score* 92,20%. Nilai *recall* yang tinggi pada strategi LP-FT juga menunjukkan kemampuan model dalam meminimalkan kesalahan diagnosis pada kasus ganas (*malignant*), yang sangat penting dalam aplikasi medis. Secara keseluruhan, strategi LP-FT memberikan hasil terbaik di kedua model, menunjukkan performa yang optimal pada metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta kemampuan generalisasi yang lebih baik pada citra medis. Namun demikian, penelitian ini memiliki keterbatasan karena hanya membandingkan dua arsitektur utama dengan tiga strategi pelatihan tertentu, sehingga kemungkinan masih terdapat kombinasi model dan metode lain yang belum dieksplorasi dan berpotensi memberikan hasil lebih optimal. Selain itu, penelitian ini belum mencakup arsitektur yang dioptimalkan untuk efisiensi komputasi, ataupun pendekatan berbasis *ensemble learning* yang dapat meningkatkan kinerja model. Untuk penelitian selanjutnya, disarankan untuk memperluas analisis dengan melibatkan lebih banyak arsitektur dan teknik *fine-tuning*, menerapkan *explainable AI* (XAI) untuk meningkatkan interpretabilitas hasil prediksi, serta menguji model pada dataset medis yang lebih bervariasi agar hasil yang diperoleh semakin representatif dan dapat diaplikasikan di dunia nyata.

REFERENCES

[1] S. Gondhowiardjo *et al.*, "Five-Year Cancer Epidemiology at the National Referral Hospital: Hospital-Based Cancer Registry Data in Indonesia," *JCO Glob. Oncol.*, no. 7, pp. 190–203, 2021, doi: 10.1200/go.20.00155.
[2] Y. L. Vechtomova, T. A. Telegina, A. A. Buglak, and M. S. Kritsky, "Uv Radiation in Dna Damage and Repair Involving Dna-photolyases and Cryptochromes," *Biomedicines*, vol. 9, no. 11, pp. 1–13, 2021, doi: 10.3390/biomedicines9111564.
[3] L. Zhou, Y. Zhong, L. Han, Y. Xie, and M. Wan, "Global, Regional, and National Trends in the Burden of Melanoma and Non-melanoma Skin Cancer: Insights From the Global Burden of Disease Study 1990–2021," *Sci. Rep.*, vol. 15, no. 1, p. 5996, Feb. 2025, doi: 10.1038/s41598-025-90485-3.
[4] L. T. Simanjuntak, "Genetic Predisposition to Malignant Melanoma in the Population of Batam, Indonesia: A Case-Control Study," *Sci. J. Dermatology Venereol.*, vol. 1, no. 2, pp. 87–99, Sep. 2023, doi: 10.59345/sjdv.v1i2.57.
[5] A. S. Al-Waisy, S. Al-Fahdawi, M. I. Khalaf, M. A. Mohammed, B. Al-Attar, and M. N. Al-Andoli, "A Deep Learning Framework for Automated Early Diagnosis and Classification of Skin Cancer Lesions in Dermoscopy Images," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-15655-9.
[6] A. Bello, S. C. Ng, and M. F. Leung, "Skin Cancer Classification Using Fine-Tuned Transfer Learning of DENSENET-121," *Appl. Sci.*, vol. 14, no. 17, 2024, doi: 10.3390/app14177707.
[7] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is Attention all You Need in Medical Image Analysis? A Review," *IEEE J. Biomed. Heal. Informatics*, vol. 28, no. 3, pp. 1398–1411, 2024, doi: 10.1109/JBHI.2023.3348436.
[8] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2021, *arXiv*. doi: 10.48550/arXiv.2010.11929.
[9] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 9992–10002, 2021, doi: 10.1109/ICCV48922.2021.00986.
[10] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 11966–11976, 2022, doi: 10.1109/CVPR52688.2022.01167.
[11] M. Ramadhan and J. Zeniarja, "Implementation of Deep Transfer Learning and Explainable AI in Skin Cancer Classification," *SISTEMASI*, vol. 14, p. 2266, Sep. 2025, doi: 10.32520/stmsi.v14i5.5425.
[12] A. Newaz, A. U. R. Adib, R. Sahil, and M. Mehzad, "An End-to-End Deep Learning Framework for Arsenicosis Diagnosis Using Mobile-Captured Skin Images," 2025, *arXiv*. doi: 10.48550/arXiv.2509.08780.
[13] G. H. Dagnaw, M. El Mouhtadi, and M. Mustapha, "Skin Cancer Classification Using Vision Transformers and Explainable Artificial Intelligence," *J. Med. Artif. Intell.*, vol. 7, no. January, pp. 14–14, Jun. 2024, doi: 10.21037/jmai-24-6.
[14] E. A. Taufik, A. Khondoker, A. F. Parsa, and S. A. M. Mostafa, "Visual Bias and Interpretability in Deep Learning for Dermatological Image Analysis," in *2025 4th International Conference on Image Processing and Media Computing (ICIPMC)*, IEEE, Jun. 2025, pp. 45–49. doi: 10.1109/ICIPMC66319.2025.11170678.
[15] J. Vieira, F. Mendonça, and F. Morgado-Dias, "Deep Learning Approaches for Skin Lesion Detection," *Electronics*, vol. 14, no. 14, p. 2785, 2025, doi: 10.3390/electronics14142785.
[16] A. Davila, J. Colan, and Y. Hasegawa, "Comparison of Fine-tuning Strategies for Transfer Learning in Medical Image Classification," *Image Vis. Comput.*, vol. 146, p. 105012, Jun. 2024, doi: 10.1016/j.imavis.2024.105012.
[17] A. Kumar, A. Raghunathan, R. Jones, T. Ma, and P. Liang, "Fine-Tuning Can Distort Pretrained Features and Underperform Out-of-Distribution," *ICLR 2022 - 10th Int. Conf. Learn. Represent.*, vol. 10, no. Id, pp. 1–54, 2022, doi: 10.48550/arXiv.2202.10054.
[18] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2240−2249
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9015

image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.

[19] L. Alzubaidi *et al.*, "Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data," *Cancers (Basel).*, vol. 13, no. 7, p. 1590, Mar. 2021, doi: 10.3390/cancers13071590.

[20] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, pp. 4023–4031, 2024, doi: 10.53555/AJBR.v27i4S.4345.

[21] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.

[22] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Iriananda, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, 2024, doi: 10.62411/faith.2024-11.

[23] R. Barinov, V. Gai, G. Kuznetsov, and V. Golubenko, "Automatic Evaluation of Neural Network Training Results," *Computers*, vol. 12, no. 2, p. 26, 2023, doi: 10.3390/computers12020026.