

Implementasi *Deep Transfer Learning* dan *Explainable AI* dalam Klasifikasi Kanker Kulit

Implementation of Deep Transfer Learning and Explainable AI in Skin Cancer Classification

^{1,2}Muhammad Eky Ramadhan*, ²Junta Zeniarja

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro di Kota Semarang, Indonesia

*e-mail: ekyramadhan054@gmail.com

(received: 17 June 2025, revised: 30 June 2025, accepted: 30 June 2025)

Abstrak

Kanker kulit merupakan salah satu penyakit yang prevalensinya terus meningkat, khususnya di wilayah dengan paparan sinar ultraviolet (UV) tinggi. Tantangan utama diagnosis kanker kulit terletak pada kemiripan visual antara lesi jinak dan ganas, menyebabkan tingginya tingkat kesalahan diagnosis, bahkan oleh tenaga medis berpengalaman. Pendekatan diagnosis berbasis kecerdasan buatan (AI) saat ini masih dianggap sebagai sistem “black box” karena kurangnya transparansi dalam proses pengambilan keputusan. Studi ini bertujuan untuk mengembangkan sistem klasifikasi kanker kulit berbasis *deep transfer learning* yang terintegrasi dengan *Explainable Artificial Intelligence* (XAI). Model *pretrained CNN*, ResNet101 dan DenseNet201 diterapkan untuk klasifikasi citra dari dataset ISIC2020 dalam dua kelas, yaitu *benign* dan *malignant*. Evaluasi kinerja dilakukan menggunakan *accuracy*, *precision*, *recall*, F1-score, *confusion matrix*, dan kurva ROC. Hasil penelitian menunjukkan bahwa ResNet101 memberikan performa klasifikasi terbaik dengan *accuracy* 87,00%, AUC 0,943, dan *recall* 89,6% dalam mendeteksi *malignant*. Untuk mendukung interpretabilitas, teknik Grad-CAM dan LIME diterapkan untuk visualisasi area penting dalam citra yang menjadi dasar keputusan model. Temuan ini menekankan pendekatan yang menjanjikan untuk mendukung sistem pendukung keputusan medis yang lebih akurat, transparan, dan dapat dijelaskan, khususnya dalam konteks diagnosis kanker kulit.

Kata kunci: klasifikasi citra, kanker kulit, transfer learning, explainable AI

Abstract

Skin cancer is one of the most prevalent diseases, especially in regions with high ultraviolet (UV) light exposure. The main challenge of skin cancer diagnosis lies in the visual similarity between benign and malignant lesions, leading to a high rate of misdiagnosis, even by experienced medical personnel. Current artificial intelligence (AI)-based diagnosis approaches are still considered as “black box” systems due to the lack of transparency in the decision-making process. This study aims to develop a skin cancer classification system based on deep transfer learning integrated with Explainable Artificial Intelligence (XAI). Pretrained CNN, ResNet101 and DenseNet201 models are applied for image classification from ISIC dataset in two classes, namely benign and malignant. Performance evaluation was performed using accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. The results showed that ResNet101 provided the best classification performance with an accuracy of 87.00%, AUC of 0.943, and recall of 89.6% in detecting malignant. To support interpretability, Grad-CAM and LIME techniques were applied to visualize important areas in the image on which the model's decisions were based. These findings emphasize a promising approach to support a more accurate, transparent, and explainable medical decision support system, particularly in the context of skin cancer diagnosis.

Keywords: image classification, skin cancer, transfer learning, explainable AI

1 Pendahuluan

Kulit merupakan organ terluar pada tubuh manusia yang berfungsi sebagai pelindung dari patogen, radiasi ultraviolet (UV), dan kehilangan cairan, sekaligus menjadi media interaksi antara tubuh dengan lingkungan [1]. Fungsi ini membuat kulit rentan terhadap berbagai gangguan, termasuk kanker kulit, yang dapat dipicu oleh faktor eksternal seperti paparan sinar UV berlebih, polusi, serta faktor internal seperti mutasi genetik dan penurunan sistem imun [2]. Secara global, kanker kulit menempati urutan kelima penyebab utama beban kesehatan, dengan prevalensi meningkat di wilayah dengan indeks UV ekstrem (10-12). Gaya hidup yang banyak melakukan aktivitas di luar ruangan tanpa proteksi memadai cenderung memiliki potensi menderita kanker kulit [3], [4].

Diagnosis dini pada awal gejala dapat meningkatkan peluang kesembuhan dan bertahan hidup pasien. Namun, saat ini diagnosis kanker kulit masih bergantung pada pemeriksaan visual, dermoskopi, dan biopsi, yang akurasi sangat dipengaruhi oleh pengalaman dokter [5]. Studi oleh Shreberk-Hassidim et al. (2023) mengungkapkan bahwa tantangan diagnosis terletak pada kemiripan antara lesi jinak seperti *dysplastic nevus* atau *seborrheic keratosis* dengan lesi ganas seperti *melanoma*. Karakteristik seperti asimetri, batas tak beraturan, dan variasi warna sering muncul pada kedua kategori. Hal ini menyebabkan tingginya tingkat kesalahan diagnosis, bahkan oleh dokter berpengalaman. Spyridonos et Al. [7] mencatat bahwa kasus biopsi awal lesi kulit dicurigai ganas ternyata bersifat jinak, mengindikasikan ketidakpastian yang signifikan dalam penilaian klinis. Tantangan serupa juga muncul dalam sistem diagnosis berbasis kecerdasan buatan (AI) yang sering bersifat *black box*, yang mana model memberikan prediksi tanpa penjelasan yang dapat diverifikasi [8]. Kurangnya transparansi ini menjadi hambatan utama adopsi AI dalam praktik klinis, mengingat tenaga medis memerlukan penjelasan logis untuk memastikan hasil keputusan model.

Seiring kemajuan teknologi, pendekatan *deep transfer learning* menawarkan solusi inovatif. Teknik ini memungkinkan pemanfaatan model *Convolutional Neural Network* (CNN) yang telah dilatih pada dataset berskala besar seperti *ImageNet*, lalu disesuaikan untuk mengenali pola spesifik terhadap citra medis, termasuk kanker kulit [9]. Berbagai studi menunjukkan bahwa pendekatan ini mampu mencapai akurasi klasifikasi kanker kulit di atas 90% [10], [11], [12], serta mampu mempercepat proses diagnosis dan menurunkan kesalahan dibandingkan metode konvensional [13]. Namun demikian, akurasi tinggi tidak selalu sejalan dengan kepercayaan pengguna. Menurut Adnan et al. [14] model prediksi masih bersifat *black box*, yaitu model menghasilkan prediksi tanpa penjelasan yang dapat divalidasi secara medis.

Untuk menjawab tantangan tersebut, pendekatan *Explainable Artificial Intelligence* (XAI) digunakan sebagai penghubung antara prediksi model dan penilaian klinis. Teknik XAI yang umum adalah *Gradient-Weighted Class Activation Mapping* (Grad-CAM). Grad-CAM menghasilkan *heatmap* yang menampilkan area lesi kritis yang diprediksi sebagai kanker kulit [15]. Selain itu, metode *Local Interpretable Model-Agnostic Explanations* (LIME) juga digunakan untuk memberikan penjelasan lokal yang berbasis model *surrogate* sederhana, seperti regresi linier, untuk menjelaskan kontribusi fitur input terhadap prediksi utama [16]. Studi oleh Khiat et al. [17] menunjukkan bahwa kombinasi Grad-CAM dan LIME menghasilkan visualisasi yang dinilai bermanfaat oleh tenaga medis karena menggabungkan informasi spasial dan numerik yang saling melengkapi.

Berdasarkan latar belakang tersebut, studi ini bertujuan untuk mengeksplorasi model klasifikasi penyakit kulit berbasis *deep transfer learning* yang transparan melalui penerapan XAI, dengan fokus pada klasifikasi kanker kulit dari *dataset* citra dermoskopi. Pertama, kinerja model *pre-trained* ResNet101 dan DenseNet201 dievaluasi dalam membedakan *benign* dan *malignant*. Kedua, menerapkan dan mengintegrasikan Grad-CAM dan LIME untuk memberikan visualisasi yang dapat dijelaskan dasar pengambilan keputusan model. Diharapkan, pendekatan ini tidak hanya mencapai performa akurasi yang tinggi, tetapi juga mampu memberikan interpretasi yang valid. Dengan demikian, studi ini berpotensi menjadi langkah transformatif dalam meningkatkan kecepatan diagnosis, mengurangi kesalahan, dan memperkuat kepercayaan tenaga medis terhadap adopsi sistem AI dalam praktik klinis.

2 Tinjauan Literatur

Kemajuan dalam bidang *deep learning* telah mendorong transformasi signifikan dalam sistem diagnosis medis berbasis citra, termasuk dalam tugas klasifikasi kanker kulit. Salah satu pendekatan

<http://sistemasi.ftik.unisi.ac.id>

yang menonjol adalah *deep transfer learning*, yaitu pemanfaatan model *convolutional neural network* (CNN) yang telah dilatih pada dataset berskala besar, seperti ImageNet, kemudian disesuaikan kembali untuk tugas klasifikasi khusus, seperti identifikasi lesi kulit jinak (*benign*) dan ganas (*malignant*).

Sejumlah studi menunjukkan bahwa pendekatan ini tidak hanya mempercepat proses pelatihan model, tetapi juga menghasilkan akurasi tinggi dalam pengenalan pola pada citra dermoskopi. Misalnya, Jain et al. [18] mengevaluasi performa enam arsitektur *transfer learning*, yaitu VGG19, InceptionV3, InceptionResNetV2, ResNet50, Xception, dan MobileNet, menggunakan dataset HAM10000 untuk klasifikasi multi-kelas lesi kulit. Hasil evaluasi menunjukkan bahwa model Xception mencapai akurasi tertinggi sebesar 90,48% serta menunjukkan performa terbaik dari segi metrik *recall* dan F1-score. Keberhasilan ini memperkuat bahwa arsitektur *deep learning* yang dikembangkan untuk klasifikasi umum dapat secara efektif ditransfer ke dalam bidang medis.

Di sisi lain, studi oleh Mateen et al. [19] mengusulkan pendekatan hibrida dalam klasifikasi melanoma. Dalam studinya, U-Net digunakan untuk segmentasi awal, Inception-ResNet-v2 untuk *feature extraction*, serta Vision Transformer untuk klasifikasi. Model ini menghasilkan akurasi hingga 98,65% dan *recall* 99,20% pada dataset ISIC2020. Pendekatan integratif seperti ini menunjukkan potensi besar dalam meningkatkan akurasi deteksi melanoma dengan menggabungkan kekuatan segmentasi spasial dan analisis global fitur.

Selanjutnya, Thwin & Park [20] mengembangkan pendekatan *ensemble learning* untuk meningkatkan ketahanan klasifikasi terhadap variabilitas citra. Tiga model arsitektur CNN populer, yaitu VGG16, InceptionV3, dan ResNet50, digabungkan dengan model segmentasi seperti U-Net, SegNet, dan DeepLabV3. Melalui strategi *oversampling* pada dataset ISIC2018, sistem ini berhasil mencapai *accuracy* 97% dan menunjukkan stabilitas terhadap variasi ukuran dan bentuk lesi. Meskipun demikian, sebagian besar studi tersebut menitikberatkan pada peningkatan metrik performa, tanpa mengeksplorasi aspek interpretabilitas dan transparansi model.

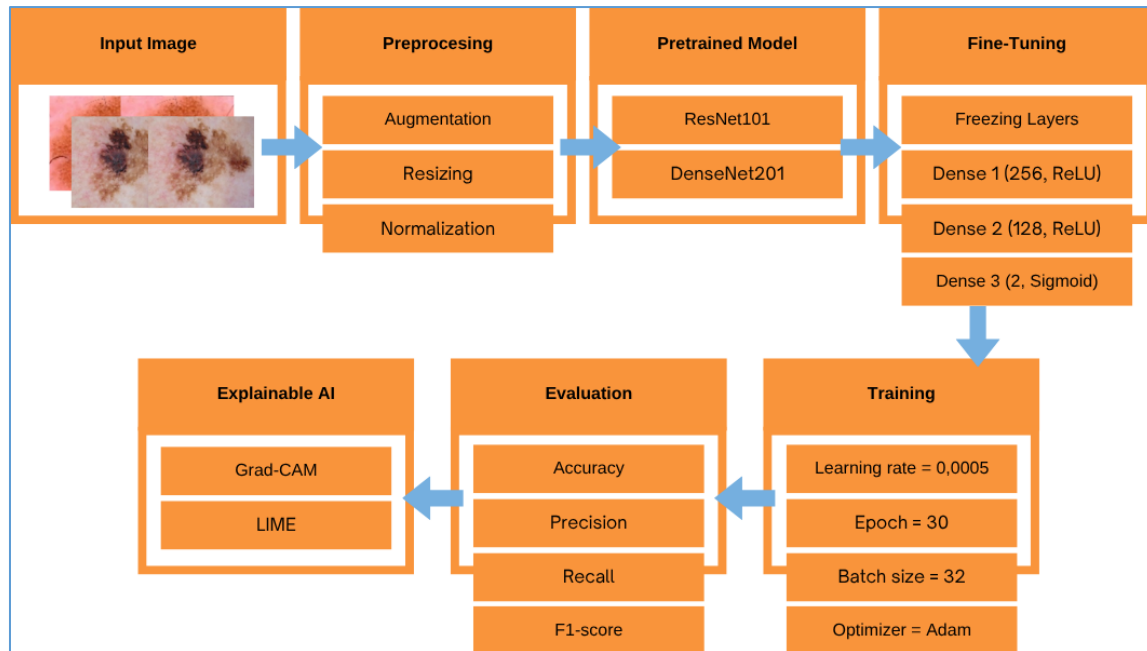
Kinerja klasifikasi yang tinggi belum tentu menjamin adopsi sistem AI dalam lingkungan medis. Tantangan utama terletak pada keterbatasan interpretabilitas, yang menjadi perhatian utama dalam sistem diagnosis berbasis AI. Alrabai et al. [21] menekankan pentingnya dasar keputusan model melalui integrasi XAI. Dalam studinya, teknik XAI diimplementasikan pada model ResNet50 dan Vision Transformer (ViT) untuk menjelaskan fitur penting yang berkontribusi terhadap klasifikasi.

Studi oleh Khan et al. [22] mengusulkan kerangka kerja berbasis *deep learning* dan *explainable AI* untuk klasifikasi COVID-19 dari citra X-ray dada. Dalam pendekatan ini, dua model *deep learning* yang telah dimodifikasi dilatih menggunakan *transfer learning*, dan fitur yang dihasilkan dari kedua model digabungkan menggunakan *discriminant canonical correlation analysis* (DCCA). Proses optimisasi fitur dilakukan dengan algoritma hibrida Whale-Elephant Herding, dan klasifikasi akhir menggunakan *Extreme Learning Machine* (ELM). Selain itu, teknik Grad-CAM diterapkan untuk menghasilkan visualisasi spasial yang menjelaskan fokus perhatian model terhadap area penting dalam citra. Hasil eksperimen menunjukkan akurasi tinggi, hingga 99,1% dan peningkatan interpretabilitas, yang memperkuat relevansi integrasi XAI dalam sistem diagnosis berbasis AI, khususnya pada aplikasi medis yang kritis.

Berdasarkan analisis terhadap berbagai studi terdahulu, dapat disimpulkan bahwa meskipun berbagai pendekatan *deep learning* telah berhasil mencapai performa klasifikasi yang tinggi, studi-studi sebelumnya cenderung berfokus pada optimalisasi performa metrik tanpa mengevaluasi integrasi XAI secara menyeluruh dan sistematis. Studi ini hadir untuk mengisi celah tersebut dengan mengusulkan implementasi model ResNet101 dan DenseNet201 berbasis *transfer learning* untuk klasifikasi biner kanker kulit, serta menerapkan teknik XAI Grad-CAM dan LIME guna memberikan visualisasi spasial atas area citra yang berkontribusi terhadap prediksi model.

3 Metode Penelitian

Studi ini menggunakan pendekatan eksperimental *deep transfer learning* untuk klasifikasi dua kelas citra lesi kulit, yaitu *benign* dan *malignant*. Studi ini juga mengintegrasikan pendekatan XAI guna meningkatkan interpretabilitas model klasifikasi *pretrained* berbasis CNN, yaitu ResNet101 dan DenseNet201. Metode penelitian terstruktur diilustrasikan pada Gambar 1.

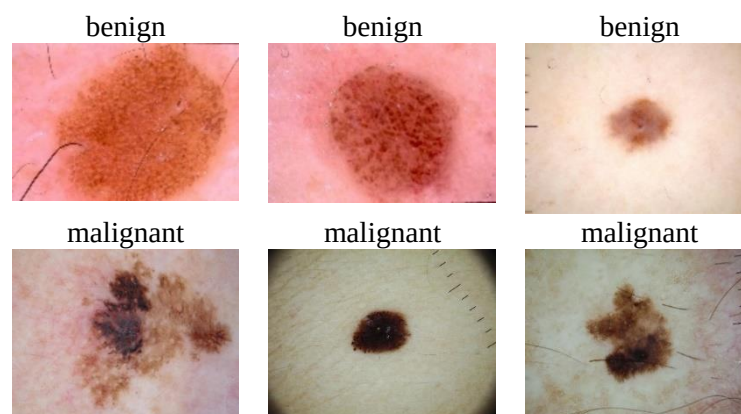


Gambar 1. Gambaran umum metode yang diusulkan

Metode penelitian yang ditunjukkan diawali dengan pengumpulan data citra lesi kulit, kemudian data diolah dengan teknik *preprocessing* data yang ada untuk memastikan data siap digunakan. Studi ini mengevaluasi kinerja model ResNet101 dan DenseNet201 dalam klasifikasi biner kanker kulit. Untuk memberikan transparansi klasifikasi model, studi ini mengintegrasikan teknik XAI, yaitu Grad-CAM dan LIME. Teknik ini diterapkan pada salah satu model yang mendapatkan performa terbaik selama pelatihan.

1. Dataset

Studi ini menggunakan *dataset* publik yang diperoleh dari ISIC Archive. *Dataset* terdiri dari 7,303 citra dermoskopi yang terbagi ke dalam dua kelas, dengan 3,376 citra *benign* dan 3,927 citra *malignant*. Data dibagi menjadi subset pelatihan dan pengujian dengan menjaga distribusi kelas untuk menghindari bias.



Gambar 2. Sampel citra yang digunakan

Untuk memberikan gambaran mengenai *dataset* yang digunakan pada studi ini, Gambar 2 menampilkan sampel gambar dari kedua kelas, yaitu *benign* dan *malignant*. Untuk memaksimalkan performa model, data pelatihan dibagi dengan perbandingan 80:20. Di mana 80% dari total citra digunakan untuk pelatihan dan 20% digunakan untuk validasi.

Tabel 1. Deskripsi dataset

	Training	Validation	Testing	Total
Benign	2,121	676	579	3,376
Malignant	2,551	786	590	3,927

Tabel 1 mendeskripsikan *dataset* yang digunakan pada studi ini. Pembagian data ini penting untuk memastikan bahwa model yang dibangun dapat di-*generalize* dengan baik pada data yang belum pernah dilihat sebelumnya [23].

2. Prapemrosesan Data

Tahap *preprocessing* merupakan langkah penting dalam *machine learning* untuk mempersiapkan citra agar dapat digunakan secara efektif, sehingga model mendapatkan hasil klasifikasi yang optimal. *Data preprocessing* mencakup tugas-tugas seperti menghilangkan *outliers*, menangani *missing values*, dan *normalization*. *Data augmentation* merupakan teknik yang digunakan untuk memperluas kumpulan data. *Data augmentation* penting dilakukan untuk menghasilkan lebih banyak data dan mengurangi risiko *overfitting*. Teknik *data augmentation* yang efektif diterapkan pada data citra meliputi transformasi geometris, mencakup *rotation*, *scaling*, *translation*, dan *flipping*, tanpa mengubah label kelasnya [24].

Pada studi ini, tahap *prerprocessing* meliputi *resizing* dan *data augmentation*. Ukuran citra diubah menjadi 224 x 224 x 3 untuk menormalkan nilai pikselnya. Ukuran ini disesuaikan dengan model *pretrained* CNN yang digunakan, di mana model ResNet101 dan DenseNet201 memerlukan *input* dengan ukuran citra yang konsisten. *Data Augmentation* diterapkan untuk meningkatkan keragaman data pelatihan dan membantu model mengatasi *overfitting*. Teknik *data augmentation* yang digunakan terdiri dari serangkaian transformasi termasuk *translation*, *rotation*, *zoom*, *horizontal flip* dan *vertical flip*, dan *brightness adjustment*.

3. Transfer Learning dan Fine Tuning

Pada tahap *feature extraction*, studi ini menerapkan metode *deep transfer learning* untuk memanfaatkan keunggulan dari model *pretrained* yang telah dilatih pada *dataset* berskala besar, seperti *ImageNet*. Strategi ini dilakukan untuk mengatasi keterbatasan jumlah data pelatihan yang umum ditemukan di bidang medis. Model yang dipilih sebagai *feature extractor* yaitu ResNet101 dan DenseNet201. Model-model ini dipilih berdasarkan kinerjanya dalam tugas klasifikasi citra medis dan potensinya untuk meningkatkan interpretabilitas dalam klasifikasi kanker kulit secara otomatis

a. ResNet101

ResNet101 dipilih karena kedalaman arsitekturnya yang mampu menangkap fitur kompleks dalam citra medis melalui 101 *convolutional layers*. Model ini mengatasi masalah *vanishing gradient* dengan mekanisme *residual connections* [25] yang mempertahankan informasi selama *gradient propagation*. ResNet101 berfokus pada kapasitas *feature extraction* hierarkis yang lebih kaya, mulai dari tepi sederhana hingga pola tekstur kompleks pada lesi kulit. Kedalaman ini memungkinkan identifikasi fitur halus seperti pigmentasi mikroskopis, batas lesi tidak teratur, dan varian warna yang kritis dalam diagnosis.

b. DenseNet201

DenseNet201 dipilih karena arsitektur uniknya yang mengadopsi *dense connectivity* antar *convolutional layers*. Setiap lapisan dalam DenseNet201 terhubung ke semua lapisan berikutnya dalam blok *dense block*, memungkinkan aliran gradien yang lebih baik dan *feature reuse* secara maksimal. Karakteristik ini membuat DenseNet201 efektif dalam menangkap variasi visual kompleks seperti pigmentasi *non-uniform*, transisi warna *subtle* pada *melanoma early-stage*, atau pola tekstur *reticular* karsinoma sel basal.

c. Adaptasi Arsitektur dan Penyesuaian Model

Pelatihan model dilakukan dalam dua fase, yaitu *transfer learning* dan *fine-tuning*. Pada fase *transfer learning*, seluruh *convolutional layers* dari model *pretrained* ResNet101 dan DenseNet201 dibekukan (*freezing layers*) untuk mempertahankan fitur-fitur umum yang telah dipelajari dari *dataset* ImageNet, terutama fitur dasar yang bersifat universal dalam pengolahan citra. Setelah bagian dasar model dibekukan, ditambahkan *classification head* baru di atasnya. Pertama, diterapkan GlobalAveragePooling2D untuk meratakan hasil *feature extraction*. Selanjutnya, dua *dense layers* dengan aktivasi ReLu ditambahkan, masing-masing dilengkapi dengan *L2 regularization* untuk mencegah *overfitting*. Terakhir, model ditutup dengan *dense output layers* berisi dua *neuron* dan menggunakan fungsi aktivasi *sigmoid* untuk memungkinkan klasifikasi biner antara *benign* dan *malignant* berbasis probabilitas. Setelah pelatihan awal selesai,

<http://sistemasi.ftik.unisi.ac.id>

tahap *fine-tuning* dijalankan dengan membuka sebagian *convolutional layers* akhir agar model dapat disesuaikan terhadap pola dan distribusi data spesifik dari *dataset* kanker kulit yang digunakan.

4. Implementasi dan Evaluasi

Dua model yang telah dilatih sebelumnya yaitu ResNet101 dan DenseNet201 digunakan untuk klasifikasi kanker kulit. Model yang digunakan disesuaikan untuk tugas -tugas tertentu, khususnya klasifikasi biner. Berbagai teknik *hyperparameter* dan *fine-tuning* diterapkan, kinerja model dipantau untuk memastikan model mendapatkan akurasi yang optimal untuk klasifikasi kanker kulit jinak dan ganas. Proses pelatihan menggunakan *Adam optimizer*, dengan *learning rate* awal sebesar 0,0005 dan *batch size* sebanyak 32. Seluruh model dilatih selama 30 *epoch* untuk mencapai kestabilan performa.

Evaluasi kinerja dilakukan dengan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan F1-score, seperti pada formula berikut:

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

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

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

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

TP menunjukkan *true positive*, *TN* menunjukkan *true negative*, *FP* menunjukkan *false positive*, *FN* menunjukkan *false negative*.

5. Integrasi XAI

Untuk meningkatkan transparansi tentang proses klasifikasi dan keputusan prediksi model, studi ini mengintegrasikan dua teknik XAI, yaitu Grad-CAM dan LIME. Kedua metode ini digunakan untuk memberikan interpretasi visual maupun numerik atas keluaran model *deep learning*, sehingga mendukung validasi klinis dan meningkatkan kepercayaan pengguna terhadap sistem.

a. Grad-CAM

Metode *Gradient-weighted Class Activation* (Grad-CAM) digunakan untuk mengidentifikasi fitur spasial penting yang memengaruhi keputusan klasifikasi model. Grad-CAM bekerja dengan menghitung gradien dari kelas target terhadap aktivasi pada *convolutional layers* terakhir untuk menghasilkan *heatmap*. Peta ini menyoroti area dalam citra yang paling berkontribusi terhadap keputusan model. Dalam konteks klasifikasi kanker kulit, visualisasi Grad-CAM memungkinkan ahli medis untuk mengevaluasi apakah fokus perhatian model berada pada area lesi yang relevan atau tidak. Dengan demikian, Grad-CAM memberikan penjelasan visual yang intuitif terhadap hasil prediksi, serta meningkatkan kepercayaan terhadap sistem klasifikasi otomatis.

b. LIME

Selain visualisasi spasial, studi ini juga menerapkan metode *Local Interpretable Model-Agnostic Explanations* (LIME) untuk memberikan penjelasan numerik yang bersifat lokal terhadap hasil prediksi. LIME bekerja dengan membuat sejumlah variasi dari input asli dan mengevaluasi prediksi model terhadap data tersebut menggunakan model *surrogate* yang lebih sederhana, seperti *linear regression*. Teknik ini menghasilkan skor kontribusi masing-masing fitur input terhadap keluaran prediksi. Dalam konteks ini, LIME digunakan untuk menjelaskan alasan di balik keputusan model dalam klasifikasi lesi sebagai *benign* atau *malignant*, dengan menunjukkan fitur-fitur dominan yang memengaruhi keputusan model secara lokal untuk setiap prediksi. Hal ini memberikan interpretasi tambahan yang tidak hanya berbasis spasial, tetapi juga berbasis nilai kontribusi numerik yang dapat dikaji secara terukur.

Integrasi Grad-CAM dan LIME dilakukan untuk menghasilkan interpretasi yang saling melengkapi, Grad-CAM memberikan konteks visual pada area citra, sedangkan LIME memberikan konteks numerik pada kontribusi fitur.

4 Hasil dan Pembahasan

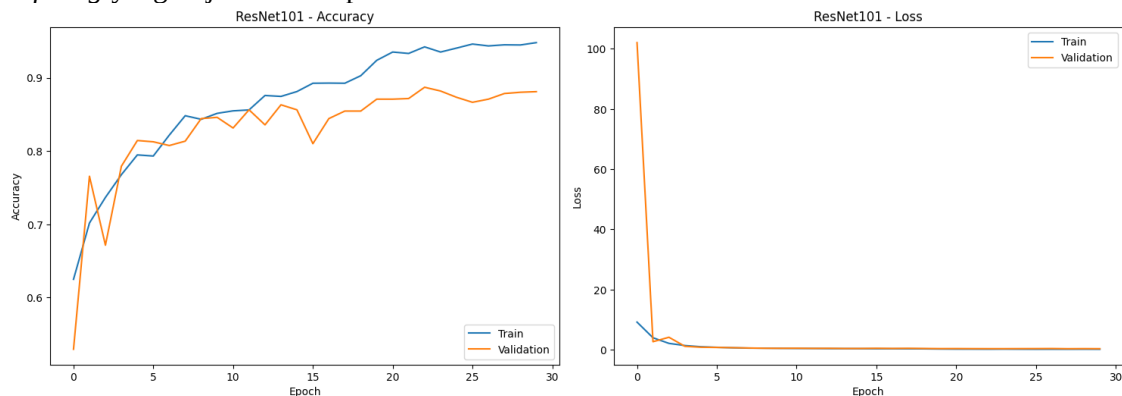
Pada studi ini, model *deep learning* diuji pada *dataset* publik ISIC untuk klasifikasi biner lesi kulit menjadi *benign* dan *malignant*. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan F1-score yang disajikan pada Table 2. Berdasarkan hasil pengujian, model ResNet101 menunjukkan performa tertinggi pada seluruh metrik yang diuji, yaitu *accuracy* 0,8700, *precision* 0,8702, *recall* 0,8700, dan F1-score 0,8698.

Tabel 2. Hasil yang diperoleh model

Model	Accuracy	Precision	Recall	F1-score
ResNet101	0,8700	0,8702	0,8700	0,8698
DenseNet201	0,8687	0,8699	0,8687	0,8688

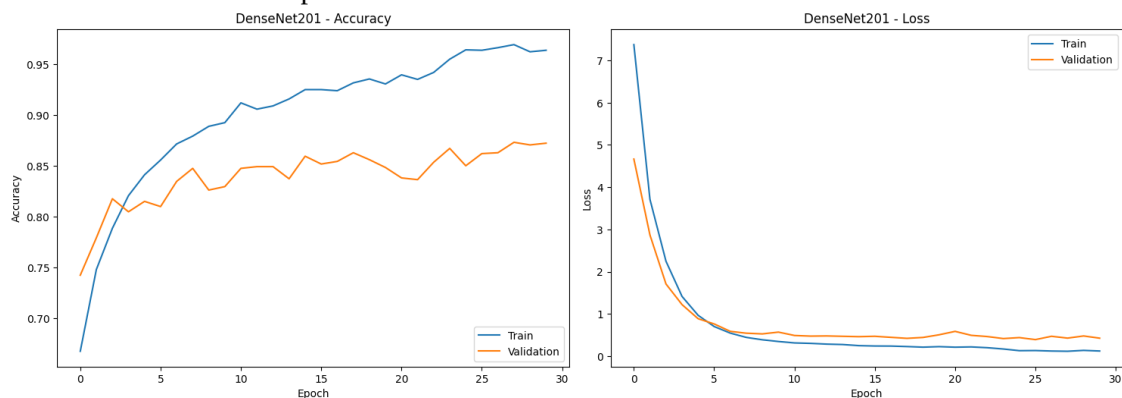
Sebaliknya, meskipun DenseNet201 juga menunjukkan performa yang tinggi, nilainya secara konsisten sedikit lebih rendah dibandingkan dengan ResNet101, terutama pada metrik *recall* dan F1-score. Dominasi ResNet101 di seluruh metrik mengindikasikan bahwa arsitektur ini memiliki kestabilan dan kapabilitas generalisasi yang lebih baik dalam mendeteksi dan membedakan pola citra lesi kulit.

Gambar 3 dan 4 menyajikan *learning curves* dalam bentuk kurva akurasi dan *loss* selama 30 *epoch* untuk masing-masing model ResNet101 dan DenseNet201. Kurva ini digunakan sebagai indikator penting dalam mengevaluasi proses pembelajaran model serta potensi *overfitting* atau *underfitting* yang terjadi selama pelatihan.



Gambar 3. Grafik nilai akurasi dan loss dengan 30 epoch pada model ResNet101

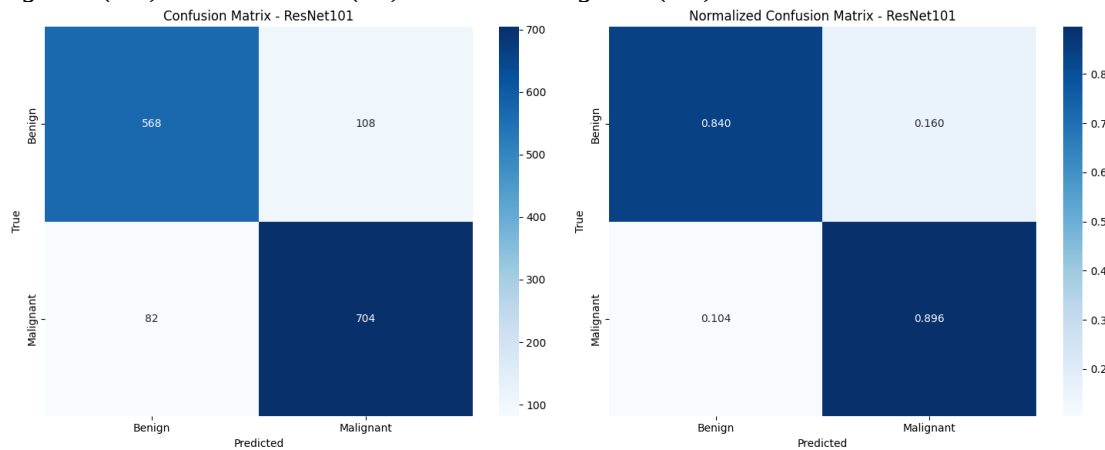
Pada grafik akurasi ResNet101 terlihat adanya tren peningkatan akurasi pada data pelatihan yang konsisten hingga mencapai nilai lebih dari 95% pada *epoch* ke-30. Akurasi validasi juga menunjukkan peningkatan yang stabil, mencapai sekitar 88% pada akhir *epoch*. Pola yang terbentuk mencerminkan proses pembelajaran yang efektif, di mana model mampu meningkatkan performanya terhadap data yang tidak terlihat secara bertahap. Di sisi lain, kurva *loss* menunjukkan penurunan tajam pada fase awal *epoch*, kemudian stabil pada nilai mendekati nol, baik untuk data pelatihan maupun validasi. Penurunan *loss* yang cepat mengindikasikan bahwa model berhasil melakukan optimisasi parameter secara efisien. Tidak terdapat indikasi *overfitting* yang signifikan karena tidak ditemukan divergensi secara mencolok antara *loss* pelatihan dan validasi.



Gambar 4. Grafik nilai akurasi dan loss dengan 30 epoch pada model DenseNet201

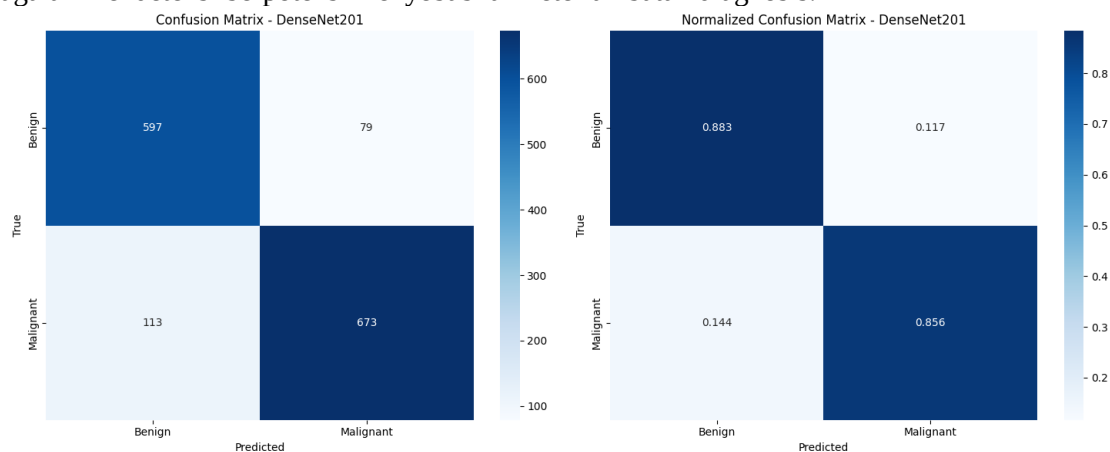
Sedangkan, pada kurva akurasi DenseNet201 menunjukkan bahwa model mengalami peningkatan performa pelatihan dari awal *epoch* hingga mencapai sekitar 96% pada akhir pelatihan. Meskipun demikian, akurasi validasi cenderung stagnan di kisaran 87% setelah *epoch* ke-10. Hal ini menunjukkan adanya potensi *overfitting* ringan, di mana model belajar terlalu spesifik terhadap data pelatihan, namun tidak sepenuhnya dapat menggeneralisasi ke data validasi. Kurva *loss* DenseNet201 juga memperlihatkan pola penurunan yang signifikan hingga mendekati konvergensi. Meskipun terdapat fluktuasi pada *loss* validasi setelah *epoch* ke-15, nilai *loss* tetap rendah secara keseluruhan yang menunjukkan bahwa model tetap stabil.

Gambar 5 dan 6 menyajikan *confusion matrix* dari model ResNet101 dan DenseNet201 dalam mengklasifikasikan kasus *benign* dan *malignant*. Evaluasi ini bertujuan untuk mengukur performa klasifikasi masing-masing model secara kuantitatif, dengan menyoroti jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).



Gambar 5. Confusion matrix pada model ResNet101

Hasil evaluasi terhadap ResNet101 menunjukkan bahwa sebesar 84,0% dari kasus *benign* berhasil diklasifikasikan secara benar sebagai *benign* (TP), sedangkan 16,0% dari kasus *benign* yang salah diklasifikasikan sebagai *malignant* (FP). Pada kategori *malignant*, model ResNet101 mampu mengklasifikasikan dengan benar sebanyak 89,6% kasus dengan tepat (TP), sementara 10,4% kasus *malignant* yang salah diklasifikasikan sebagai *benign* (FP). Angka ini menunjukkan bahwa ResNet101 memiliki *recall* tinggi dalam mendeteksi lesi ganas, meskipun dengan konsekuensi *False Positive* yang relatif lebih tinggi. Dalam konteks medis, *recall* tinggi menjadi prioritas penting karena kegagalan mendeteksi berpotensi menyebabkan keterlambatan diagnosis.

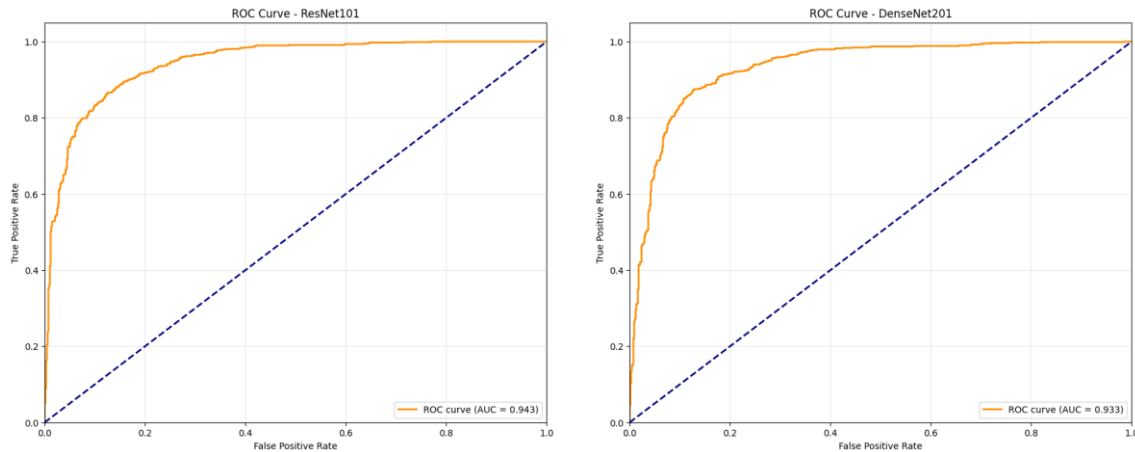


Gambar 6. Confusion matrix pada model DenseNet201

Sementara itu, hasil evaluasi terhadap DenseNet201 menunjukkan tingkat *True Negative* lebih tinggi, yaitu 88,3% dengan *False Positive* hanya sebesar 11,7%, namun tingkat *True Positive*-nya sedikit lebih rendah, yaitu 85,6%, dan *False Negative* sebesar 14,4%. Artinya, meskipun model ini lebih tepat dalam mengidentifikasi lesi jinak, terdapat risiko yang lebih besar dalam melewatkan

deteksi kanker kulit. Dengan kata lain, terdapat *trade-off* antara *recall* dan *specificity* pada kedua model. ResNet101 lebih unggul dalam mendeteksi *malignant*, sementara DenseNet201 lebih konservatif dan cenderung meminimalkan *False Positive*.

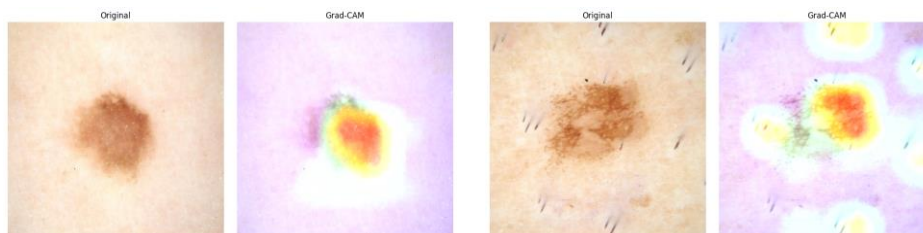
Selain evaluasi berbasis *confusion matrix*, performa model juga dianalisis menggunakan kurva *Receiver Operating Characteristic* (ROC) untuk masing-masing model, seperti yang ditampilkan pada Gambar 7. Kurva ROC merupakan representasi grafis yang menggambarkan kemampuan klasifikasi model dalam membedakan antara dua kelas, yaitu *benign* dan *malignant* pada berbagai ambang batas (*threshold*). Sumbu horizontal menunjukkan *False Positive Rate* (FPR), sedangkan sumbu vertikal menunjukkan *True Positive Rate* (TPR) atau *recall*. Salah satu metrik utama dari kurva ini adalah *Area Under the Curve* (AUC) yang mencerminkan probabilitas bahwa model akan mampu mengklasifikasikan suatu sampel positif lebih tinggi daripada sampel negatif secara acak.



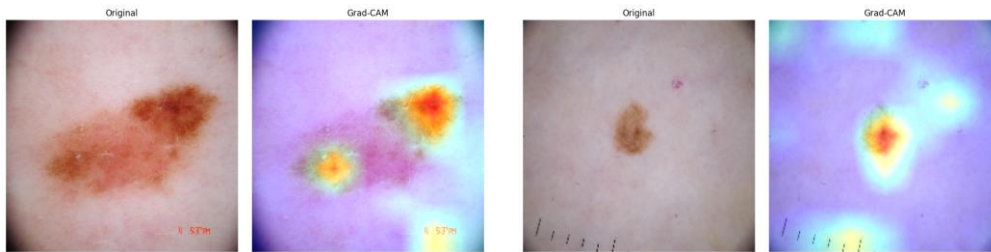
Gambar 7. Kurva ROC dari model yang diusulkan

Pada model ResNet101, diperoleh nilai AUC sebesar 0.943. Nilai ini mendekati angka maksimum 1,0 yang mengindikasikan kemampuan klasifikasi yang sangat baik. Kurva ROC ResNet101 juga memperlihatkan peningkatan tajam pada awal grafik, menandakan *recall* tinggi pada *low threshold level*. Model DenseNet201 juga menunjukkan performa yang kompetitif dengan nilai AUC sebesar 0.933. Hasil ini menunjukkan performa klasifikasi yang sangat baik, meskipun sedikit lebih rendah dibandingkan ResNet101. Kurva ROC DenseNet201 juga menunjukkan pola serupa dengan kurva ResNet101, di mana kenaikan tajam pada awal dan pendekatan terhadap titik (0,1), menandakan bahwa model memiliki *recall* dan *specificity* yang cukup seimbang. Berdasarkan analisis ROC, model ResNet101 maupun DenseNet201 menunjukkan performa yang sangat baik, dengan nilai AUC yang berada di atas 0,9. Namun, model ResNet101 sedikit lebih unggul dalam hal kemampuan diskriminatif.

Untuk meningkatkan transparansi dan interpretabilitas prediksi model, teknik XAI berupa Grad-CAM diterapkan pada model ResNet101 yang menunjukkan performa klasifikasi yang lebih unggul dibandingkan dengan DenseNet201, baik dari segi nilai *confusion matrix* maupun analisis kurva ROC. Gambar 8, 9, dan 10 menyajikan hasil penerapan teknik Grad-CAM dalam kelompok *benign* dan *malignant*. Masing-masing citra terdiri atas pasangan gambar orisinal dan peta aktivasi Grad-CAM, yang menyoroti wilayah-wilayah citra yang berkontribusi signifikan terhadap keputusan klasifikasi model.

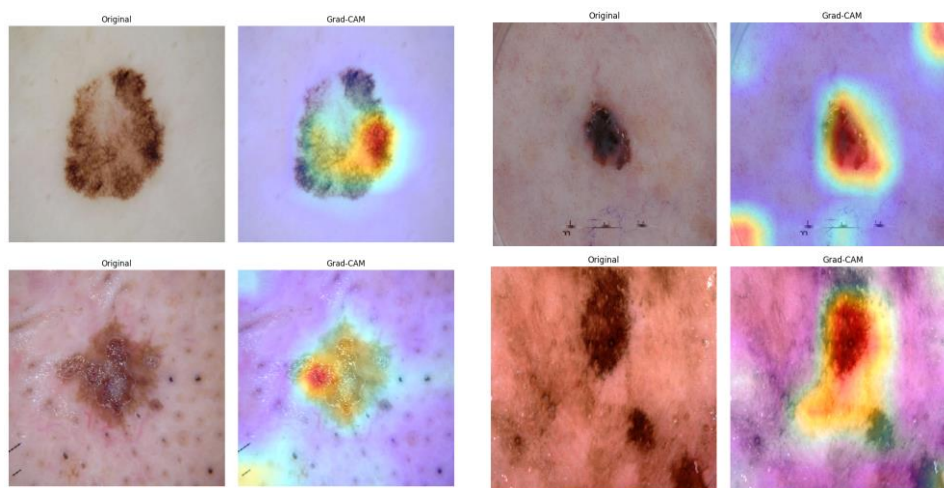


Gambar 8. Ilustrasi Grad-CAM model ResNet101 untuk sampel benign



Gambar 9. Ilustrasi Grad-CAM model ResNet101 untuk sampel benign

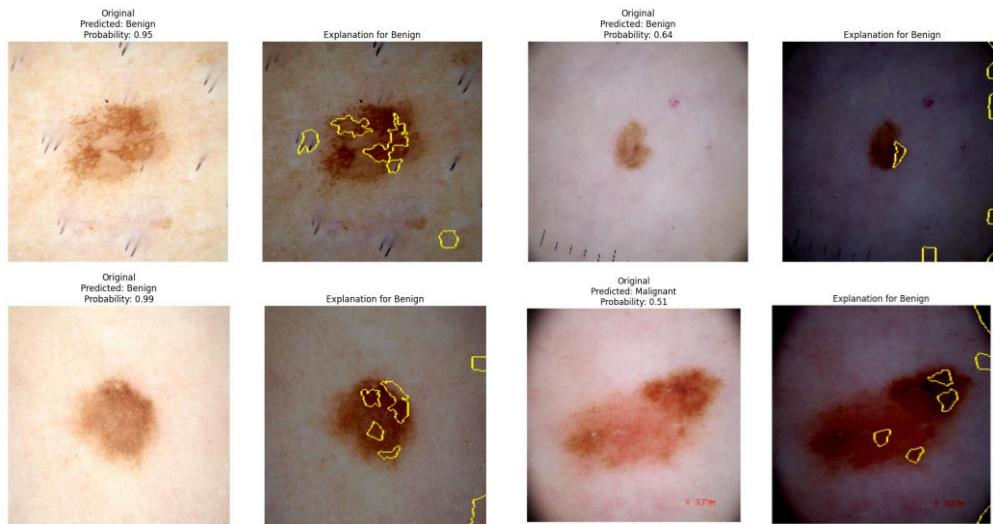
Pada Gambar 8 dan 9, terlihat bahwa Grad-CAM mampu menyoroti area inti lesi kulit *benign* secara tepat. Peta aktivasi berwarna merah hingga kuning yang muncul pada bagian tengah lesi menandakan bahwa model memberikan perhatian dominan pada struktur yang khas pada lesi *benign*, seperti batas lesi yang teratur dan warna yang homogen. Aktivasi yang terfokus ini menunjukkan bahwa ResNet101 mengandalkan fitur visual yang relevan dalam mengambil keputusan, walaupun masih terdapat bagian yang dipengaruhi oleh area di luar lesi. Visualisasi ini mengonfirmasi bahwa keputusan klasifikasi tidak diambil secara acak, tetapi berdasarkan fitur morfologis yang relevan.



Gambar 10. Ilustrasi Grad-CAM model ResNet101 untuk sampel malignant

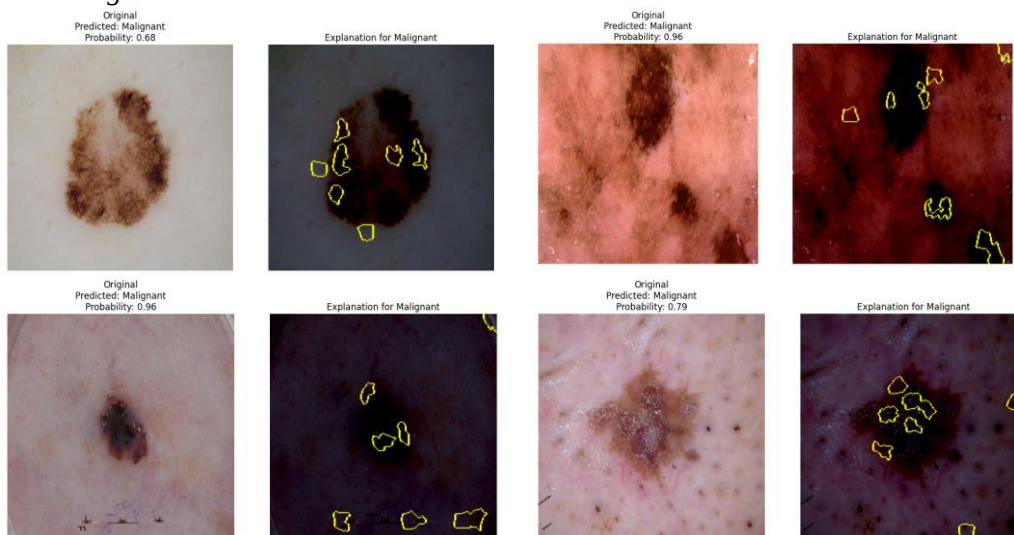
Sementara itu, Gambar 10 menunjukkan peta Grad-CAM pada dua citra *malignant* yang memperlihatkan pola aktivasi yang lebih menyebar dan intens, terutama pada area dengan tekstur tidak merata, batas lesi yang kabur, serta warna yang bervariasi. Pusat perhatian model secara konsisten tertuju pada bagian-bagian lesi yang secara klinis diasosiasikan dengan ciri-ciri melanoma atau karsinoma. Warna merah terang yang mendominasi menandakan perhatian kuat model terhadap fitur-fitur yang diasosiasikan dengan *melanoma* atau *carcinoma*, seperti asimetri, warna bervariasi, dan struktur jaringan abnormal. Pola aktivasi ini menunjukkan bahwa ResNet101 tidak hanya mengenali bentuk lesi secara keseluruhan, tetapi juga mengekstrak mikrostruktur visual penting yang membedakan antara *benign* dan *malignant*.

Selain menerapkan teknik XAI Grad-CAM untuk visualisasi *heatmap*, LIME juga digunakan untuk menjelaskan keputusan model secara lokal dan memberikan interpretasi berbasis fitur terhadap prediksi kanker kulit. Gambar 11 dan 12 menyajikan hasil penerapan LIME pada sampel dengan label prediksi *benign* dan *malignant* oleh model ResNet101. Masing-masing pasangan gambar terdiri dari citra asli di sebelah kiri dan visualisasi interpretasi LIME di sebelah kanan, di mana area yang di sorot dengan garis-garis kuning merepresentasikan bagian citra yang paling berkontribusi terhadap prediksi model.



Gambar 11. Penjelasan LIME model ResNet101 untuk sampel benign

Pada Gambar 11, citra di sebelah kiri menunjukkan lesi kulit yang diklasifikasikan sebagai *benign* beserta nilai probabilitasnya, sedangkan citra disebelah kanan memperlihatkan garis berwarna kuning yang mengelilingi wilayah-wilayah penting yang dianggap berkontribusi paling besar terhadap klasifikasi tersebut. Probabilitas tinggi menunjukkan fokus segmentasi LIME yang konsisten menyoroti area dengan ciri visual khas *benign*, seperti distribusi warna yang merata, bentuk simetris, dan batas lesi yang jelas. Hal ini menunjukkan bahwa prediksi model didasarkan pada pola yang stabil dan dikenali dengan baik. Sementara itu, pada sampel dengan keyakinan rendah, segmentasi LIME tetap menunjukkan fokus pada area yang relevan, tetapi fitur visualnya tampak lebih ambigu, yang dapat menandakan bahwa lesi tersebut memiliki karakteristik yang mendekati batas klasifikasi antara *benign* dan *malignant*.



Gambar 12. Penjelasan LIME model ResNet101 untuk sampel malignant

Sementara itu, Gambar 12 menampilkan hasil interpretasi LIME pada sampel *malignant* dengan, yang ditandai dengan tekstur kompleks, warna gelap tidak seragam, serta bentuk yang tidak simetris. Garis-garis kuning mengelilingi area penting pada bagian dalam dan tepi lesi, termasuk wilayah dengan pigmentasi ekstrem dan kontur tidak teratur. Pada nilai probabilitas rendah menunjukkan fokus yang masih terbagi pada beberapa area dengan struktur tidak terlalu jelas, menandakan bahwa model hanya sedikit lebih condong pada prediksi *malignant*. Sebaliknya, pada prediksi dengan probabilitas tinggi, visualisasi LIME menyoroti area lesi dengan pigmentasi gelap, bentuk tidak simetris, dan tekstur kompleks, ciri yang konsisten dengan lesi ganas seperti *melanoma* atau *carcinoma*.

Hasil visualisasi Grad-CAM dan LIME menunjukkan pendekatan yang saling melengkapi dalam mengungkap dasar keputusan model. Grad-CAM berhasil menyoroti area-area penting pada lesi kulit yang menjadi fokus utama model dalam mengambil keputusan klasifikasi, baik pada sampel *benign* maupun *malignant*. Sebaliknya, LIME memberikan interpretasi yang bersifat lokal dengan memvisualisasikan kontribusi fitur-fitur tertentu terhadap prediksi kelas. Namun, hasil interpretasi cenderung terfragmentasi dan hanya menyoroti sebagian area dari lesi, khususnya pada prediksi dengan nilai probabilitas rendah.

5 Kesimpulan

Studi ini menunjukkan bahwa penerapan model *deep learning* berbasis *transfer learning*, khususnya ResNet101, mampu membrikan performa klasifikasi yang sangat baik dalam membedakan lesi kulit *benign* dan *malignant* pada *dataset* ISIC. Model ResNet101 mencatatkan nilai akurasi sebesar 87.00% dan AUC sebesar 0.943, menjadikannya lebih unggul dibandingkan DenseNet201 dalam seluruh metrik evaluasi. Analisis kurva *accuracy* dan *loss* selama 30 *epoch* mengindikasikan bahwa ResNet101 belajar secara efisien tanpa menunjukkan potensi *overfitting* yang signifikan, dengan akurasi validasi yang konsisten dan stabil sejak *epoch* ke-12. Sebaliknya, meskipun model DenseNet201 menunjukkan akurasi pelatihan tinggi, akurasi validasi cenderung stagnan dan mulai menunjukkan potensi *overfitting* setelah *epoch* ke-10. Penerapan teknik XAI menggunakan Grad-CAM berhasil menyoroti area lesi kulit yang berkontribusi signifikan terhadap keputusan klasifikasi model. Hasil visualisasi ini menunjukkan bahwa keputusan model didasarkan pada fitur-fitur yang relevan, baik pada kasus *benign* dan *malignant*, sehingga meningkatkan transparansi hasil prediksi model. Di sisi lain, LIME memberikan interpretasi lokal terhadap prediksi individual, dengan menyoroti bagian-bagian citra yang secara spesifik mendukung klasifikasi *benign* maupun *malignant*. Namun, hasil interpretasi LIME pada beberapa prediksi menunjukkan keterbatasan dalam mengungkap area penting secara menyeluruh, sehingga memerlukan validasi lebih lanjut oleh ahli dermatologi untuk memastikan kesesuaiannya dengan diagnosis klinis. Dengan demikian, integrasi ResNet101 dengan Grad-CAM dan LIME memberikan pemahaman lebih komprehensif terhadap perilaku model, menggabungkan interpretasi spasial global dan lokal dalam sistem klasifikasi kanker kulit berbasis AI. Ke depan, sistem ini dapat dikembangkan untuk klasifikasi *multi-class*, memperluas teknik XAI seperti SHAP atau *Integrated Gradients*, serta diimplementasikan dalam bentuk aplikasi berbasis *web* atau *mobile* guna menunjang penggunaanya dalam praktik medis

Referensi

- [1] M. N. Alnuaimi et al., “Transfer Learning Empowered Skin Diseases Detection in Children,” *CMES - Computer Modeling in Engineering and Sciences*, Vol. 141, No. 3, pp. 2609–2623, 2024, doi: 10.32604/cmcs.2024.055303.
- [2] S. Gondhowiardjo et al., “Five-Year Cancer Epidemiology at the National Referral Hospital: Hospital-based Cancer Registry Data in Indonesia,” *JCO Global Oncol*, Vol. 7, pp. 190–203, 2021, doi: 10.1200/GO.20.
- [3] 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,” Nov. 01, 2021, MDPI. doi: 10.3390/biomedicines9111564.
- [4] R. Dandu, M. Vinayaka Murthy, and Y. B. Ravi Kumar, “Transfer learning for segmentation with hybrid classification to Detect Melanoma Skin Cancer,” *Heliyon*, Vol. 9, No. 4, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15416.
- [5] D. Aharwal, R. Changulani, A. Pathak, and D. Shukla, “Psychiatric Morbidity among Dermatological Patients Attending a Tertiary Care Hospital in Central India,” *National Journal of Community Medicine*, Vol. 14, No. 7, pp. 440–445, Jul. 2023, doi: 10.55489/njcm.140720232918.
- [6] R. Shreberk-Hassidim, S. M. Ostrowski, and D. E. Fisher, “The Complex Interplay between Nevi and Melanoma: Risk Factors and Precursors,” Feb. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ijms24043541.

- [7] P. Spyridonos, G. Gaitanis, A. Likas, and I. Bassukas, "Characterizing malignant melanoma clinically resembling seborrheic keratosis using deep knowledge transfer," *Cancers (Basel)*, Vol. 13, No. 24, Dec. 2021, doi: 10.3390/cancers13246300.
- [8] Y. Zhang, Y. Weng, and J. Lund, "Applications of Explainable Artificial Intelligence in Diagnosis and Surgery," Feb. 01, 2022, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/diagnostics12020237.
- [9] D. F. Santos-Bustos, B. M. Nguyen, and H. E. Espitia, "Towards Automated Eye Cancer Classification via VGG and ResNet Networks Using Transfer Learning," *Engineering Science and Technology, an International Journal*, Vol. 35, Nov. 2022, doi: 10.1016/j.jestch.2022.101214.
- [10] A. T. Ibrahim, M. Abdullahi, A. F. Donfack Kana, M. T. Mohammed, and I. H. Hassan, "Categorical Classification of Skin Cancer using a Weighted Ensemble of Transfer Learning with Test Time Augmentation," *Data Science and Management*, Nov. 2024, doi: 10.1016/j.dsm.2024.10.002.
- [11] R. O. Ogundokun et al., "Enhancing Skin Cancer Detection and Classification in Dermoscopic Images through Concatenated MobileNetV2 and Xception Models," *Bioengineering*, Vol. 10, No. 8, Aug. 2023, doi: 10.3390/bioengineering10080979.
- [12] O. E. Adebayo, B. Chatelain, D. Trucu, and R. Eftimie, "Deep Learning Approaches for the Classification of Keloid Images in the Context of Malignant and Benign Skin Disorders," *Diagnostics*, Vol. 15, No. 6, Mar. 2025, doi: 10.3390/diagnostics15060710.
- [13] P. N. Hai, N. C. Thanh, N. T. Trung, and T. T. Kien, "Transfer Learning for Disease Diagnosis from Myocardial Perfusion SPECT Imaging," *Computers, Materials and Continua*, Vol. 73, No. 3, pp. 5925–5941, 2022, doi: 10.32604/cmc.2022.031027.
- [14] K. M. Adnan et al., "Deep Learning Driven Interpretable and Informed Decision Making Model for Brain Tumour Prediction using Explainable AI," *Sci Rep*, Vol. 15, No. 1, Dec. 2025, doi: 10.1038/s41598-025-03358-0.
- [15] S. Alkhalaf et al., "Adaptive Aquila Optimizer with Explainable Artificial Intelligence-Enabled Cancer Diagnosis on Medical Imaging," *Cancers (Basel)*, Vol. 15, No. 5, Mar. 2023, doi: 10.3390/cancers15051492.
- [16] A. Gramegna and P. Giudici, "SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk," *Front Artif Intell*, Vol. 4, 2021, doi: 10.3389/frai.2021.752558.
- [17] S. Khiat, S. A. Mahmoudi, S. Stassin, L. Boukerroui, B. Senai, and S. Mahmoudi, "An Efficient Explainability of Deep Models on Medical Images," *Algorithms*, Vol. 18, No. 4, Apr. 2025, doi: 10.3390/a18040210.
- [18] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep Learning-based Transfer Learning for Classification of Skin Cancer," *Sensors*, Vol. 21, No. 23, Dec. 2021, doi: 10.3390/s21238142.
- [19] M. Mateen, S. Hayat, F. Arshad, Y. H. Gu, and M. A. Al-antari, "Hybrid Deep Learning Framework for Melanoma Diagnosis Using Dermoscopic Medical Images," *Diagnostics*, Vol. 14, No. 19, Oct. 2024, doi: 10.3390/diagnostics14192242.
- [20] S. M. Thwin and H. S. Park, "Skin Lesion Classification using a Deep Ensemble Model," *Applied Sciences (Switzerland)*, Vol. 14, No. 13, Jul. 2024, doi: 10.3390/app14135599.
- [21] A. Alrabai, A. Echtioui, and F. Kallel, "Exploring Pre-Trained Models for Skin Cancer Classification," *Applied System Innovation*, Vol. 8, No. 2, Apr. 2025, doi: 10.3390/asi8020035.
- [22] M. A. Khan et al., "COVID-19 Classification from Chest X-Ray Images: A Framework of Deep Explainable Artificial Intelligence," *Comput Intell Neurosci*, Vol. 2022, 2022, doi: 10.1155/2022/4254631.
- [23] Luqman Hakim, Z. Sari, and H. Handhajani, "Klasifikasi Citra Pigmen Kanker Kulit Menggunakan Convolutional Neural Network," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 5, No. 2, pp. 379–385, Apr. 2021, doi: 10.29207/resti.v5i2.3001.
- [24] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-Processing and Data Augmentation Techniques," *Global Transitions Proceedings*, Vol. 3, No. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltip.2022.04.020.

- [25] Y. Shu, J. Zhang, Y. Wang, and Y. Wei, "Fruit Freshness Classification and Detection Based on the ResNet-101 Network and Non-Local Attention Mechanism," *Foods*, Vol. 14, no. 11, p. 1987, Jun. 2025, doi: 10.3390/foods14111987.