

# Klasifikasi Lesi Kulit Multi-Kelas Otomatis untuk Deteksi Kanker pada Kulit

Josephine Sere Emanuela Tamba<sup>1</sup> Hasan Abdul Lathif<sup>2</sup> Diniya Fakhriza Aulia Putri Zahra<sup>3</sup>

<sup>1</sup> Program Studi Teknik Biomedik, Departemen Teknik Elektro, Fakultas Teknik, Universitas Indonesia

josephine.sere@ui.ac.id hasan.abdul@ui.ac.id dinya.fakhriza@ui.ac.id

**Abstrak** — Kanker kulit merupakan salah satu jenis kanker paling umum dan mematikan di seluruh dunia, di mana diagnosis yang akurat dan sedini mungkin menjadi kunci utama keberhasilan penanganan. Namun, proses diagnosis menghadapi tantangan signifikan akibat terbatasnya akses terhadap dokter spesialis kulit serta kesulitan dalam membedakan lesi jinak dan ganas secara visual. Proyek ini bertujuan untuk mengembangkan sebuah sistem pendukung keputusan berbasis kecerdasan buatan (AI-Powered Decision-Support Tool) untuk melakukan klasifikasi lesi kulit multi-kelas secara otomatis. Menggunakan dataset publik HAM10000 yang terdiri dari tujuh kelas lesi kulit, sistem ini membandingkan performa antara model dasar Convolutional Neural Network (CNN) Sekuensial dengan model DenseNet121 yang lebih canggih melalui pendekatan transfer learning dan fine-tuning. Teknik seperti augmentasi data dan pembobotan kelas (class weight) juga diimplementasikan untuk menangani tantangan pada data. Hasil evaluasi pada data uji menunjukkan bahwa model DenseNet121 mencapai akurasi sebesar 86%, yang secara signifikan lebih unggul dibandingkan model CNN baseline yang hanya mencapai akurasi 49%. Hasil ini menunjukkan bahwa sistem yang dikembangkan memiliki potensi besar untuk mendukung dermatologis dalam pengambilan keputusan klinis, meningkatkan kecepatan diagnosis, dan memperluas akses terhadap skrining yang andal.

**Keyword** — Klasifikasi Lesi Kulit, Convolutional Neural Network (CNN), Transfer Learning, DenseNet121

## I. PENDAHULUAN

Kanker kulit merupakan salah satu tantangan kesehatan global yang paling signifikan, dikategorikan sebagai salah satu jenis kanker yang paling umum dan mematikan di seluruh dunia [1]. Keberhasilan penanganan penyakit ini sangat bergantung pada kemampuan untuk mendeteksi secara dini dan akurat. Namun, tantangan utama dalam penanganannya terletak pada kesulitan diagnosis visual untuk membedakan secara akurat antara lesi yang jinak (benign) dan yang ganas (malignant). Kompleksitas ini, ditambah dengan kenyataan terbatasnya akses masyarakat terhadap keahlian dermatologis, seringkali berujung pada diagnosis yang terlambat (late diagnosis), sebuah kondisi yang secara langsung dapat memperburuk prognosis dan tingkat kelangsungan hidup pasien. Oleh karena itu, terdapat kebutuhan mendesak akan sebuah terobosan teknologi yang dapat membantu meningkatkan akurasi, konsistensi, dan aksesibilitas skrining awal lesi kulit.

Menjawab tantangan ini, proyek ini bertujuan untuk mengembangkan sebuah AI-Powered Decision-Support Tool yang mampu mengklasifikasikan berbagai jenis lesi kulit secara otomatis. Tujuan utamanya adalah untuk menciptakan sebuah sistem yang dapat menyelamatkan nyawa dengan mempersingkat masa tunggu diagnostik, serta mendemokratisasi layanan kesehatan dengan menyediakan alat skrining. Secara spesifik, solusi yang diusulkan

dirancang untuk mendukung para dermatologis dalam pengambilan keputusan klinis, meningkatkan kecepatan dan konsistensi diagnosis dengan mengotomatiskan proses skrining awal, serta meningkatkan akses terhadap skrining yang reliabel, terutama di wilayah yang kekurangan layanan medis.

Ruang lingkup proyek ini terpusat pada pemanfaatan dataset publik HAM10000 (Human Against Machine with 10000 training images). Fokus utama sistem adalah melakukan klasifikasi pada tujuh jenis lesi kulit yang berbeda. Pendekatan metodologis yang digunakan adalah dengan melatih dan membandingkan dua arsitektur model secara paralel yaitu CNN sederhana sebagai baseline dan model DenseNet121 yang diimplementasikan menggunakan strategi transfer learning dan fine-tuning. Meskipun demikian, penting untuk diakui bahwa proyek ini memiliki batasan, termasuk ketergantungan pada dataset tunggal dengan ketidakseimbangan kelas yang signifikan dan kurangnya diversitas warna kulit, serta fakta bahwa sistem ini berfungsi sebagai alat bantu keputusan dan bukan merupakan pengganti diagnosis medis profesional.

## II. METODOLOGI

### A. Dataset

Dataset yang digunakan adalah gambar-gambar dari HAM10000 berisi berbagai jenis lesi kulit [2], dan digunakan oleh model untuk melakukan klasifikasi gambar tersebut menjadi 7 jenis kulit yang berbeda, yaitu melanocytic nevi (nv), melanoma (mel), benign keratosis-like (bkl), basal cell carcinoma (bcc), actinic keratosis (aikc), vascular lesions, dan dermatofibroma (df), yang ditunjukkan pada gambar 1 [3].



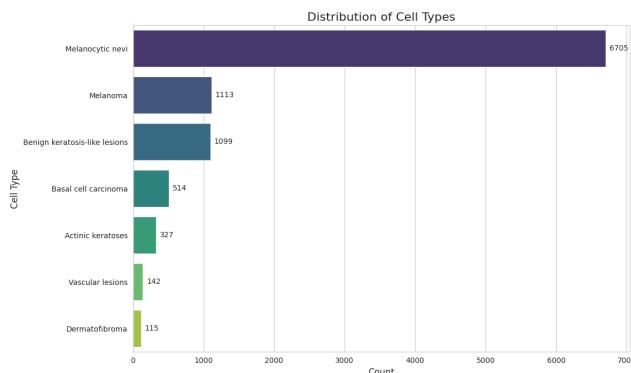
Gambar 1. Gambar Jenis Lesi Kulit

Selain gambar, dataset juga dilengkapi dengan file metadata yang berisi informasi tambahan mengenai tiap gambar pada

dataset, terdiri atas `image_id` sebagai *identifier* unik tiap gambar, `dx` sebagai diagnosa dari dermatologis, `age` sebagai umur pasien, `sex` sebagai gender pasien, dan `localization` sebagai letak dari lesi kulit dalam tubuh pasien.

## B. Exploratory Data Analysis

EDA dilakukan untuk memahami struktur, pola, dan anomali yang ada pada gambar dataset yang digunakan [4]. Proses EDA meliputi melakukan *data cleaning* untuk membersihkan *missing values*, duplikat gambar, dan gambar yang tidak konsisten. Selain itu, dilakukan plot grafik distribusi tipe lesi kulit, umur pasien, dan visualisasi gambar dari tiap lesi kulit.



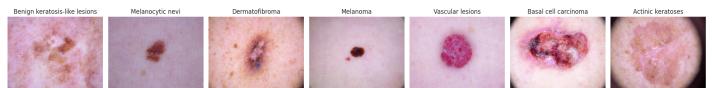
Gambar 2. Distribusi Tipe Lesi Kulit

Gambar 2 menunjukkan distribusi dari tipe lesi kulit, dan dapat dilihat bahwa lesi *melanocytic nevi* memiliki jumlah gambar yang paling banyak (6705 gambar), dan *dermatofibroma* memiliki jumlah gambar paling sedikit pada dataset (115 gambar). Distribusi ini menunjukkan dataset yang *imbalance* atau tidak seimbang, sehingga pada pembangunan model ML akan ditambahkan *weights* untuk mengatasi ketidakseimbangan jumlah data.



Gambar 3. Distribusi Umur Pasien

Gambar 3 menunjukkan distribusi umur pasien, menghasilkan penderita lesi kulit rata-rata berumur 40 tahun, dengan penderita paling sedikit terdapat pada umur  $\leq 20$  tahun. Gambar 4 merupakan visualisasi dari tiap tipe lesi kulit, menunjukkan penyakit *melanocytic nevi* dan melanoma memiliki pola yang mirip. Sehingga, dibutuhkan teknologi untuk mengklasifikasi jenis penyakit ini dengan hasil yang akurat.



Gambar 4. Visualisasi Gambar Lesi Kulit

## C. Data Splitting

Digunakan untuk membagi dataset menjadi subset-subset kecil yang berfungsi untuk melatih dan mengevaluasi model yang digunakan, dataset dibagi menjadi 3 subset, yaitu 70% *training set* untuk melatih model agar model mempelajari pola dan fitur data, 15% *validation set* untuk melakukan *hyperparameters tuning* yang menghasilkan *advanced results*, dan 15% *test set* untuk membantu performa final dari model setelah model dilatih dengan *validation* dan *test set*.

## D. Data Augmentation

Augmentasi data bertujuan untuk menghasilkan data tambahan yang berasal dari transformasi atau perubahan dataset yang sudah ada. Digunakan *ImageDataGenerator* yang akan melakukan transformasi termasuk merotasi, *zoom in/zoom out*, memindahkan, dan *flipping* gambar secara vertikal atau *horizontal*. Dengan ini, model dapat berlatih lebih baik menggunakan data yang lebih jelas.

## E. Model Architecture

Sebelum membangun arsitektur model yang digunakan, ditambahkan *class weight* untuk mengatasi ketidakseimbangan data, *weights* yang digunakan dihitung berdasarkan frekuensi dari tiap jenis lesi kulit pada *training set*. Jenis lesi kulit mayoritas memiliki *weights* lebih banyak dari lesi kulit minoritas, agar model tidak terlalu fokus pada minoritas dan meningkatkan fokus pada mayoritas [5].

Digunakan 2 jenis arsitektur model, yaitu *baseline* model menggunakan CNN dan DenseNet121 sebagai *advanced model*. Model *baseline* terdiri atas lapisan-lapisan:

- Lapisan *convolutional* dengan ReLU *activation* untuk ekstraksi fitur gambar.
- Lapisan *MaxPooling* untuk mengurangi ukuran spasial saat ekstraksi fitur, mengurangi beban komputasional saat melatih model.
- Lapisan *Dropout* untuk mencegah *overfitting* data dengan secara acak mengatur fraksi sebesar 25% selama proses latihan.
- Lapisan *flatten* untuk mengubah bentuk 2D data dari layer sebelumnya menjadi 1D.
- Lapisan *Dense* sebagai lapisan yang terhubung satu sama lain dan membantu menghasilkan prediksi final klasifikasi lesi kulit.
- Lapisan keluaran menggunakan aktivasi *softmax* untuk mengklasifikasi hasil 7 jenis lesi kulit.

Untuk model *baseline* digunakan 30 *epochs*, optimisasi Adam dan *learning rate* 0.001. DenseNet121 digunakan sebagai model untuk menghasilkan *advanced results* karena keunggulannya sebagai *pre-trained model* yang sudah dilatih

sebelumnya dengan dataset gambar-gambar lainnya sehingga sudah mengetahui berbagai fitur gambar. Lapisannya terdiri dari:

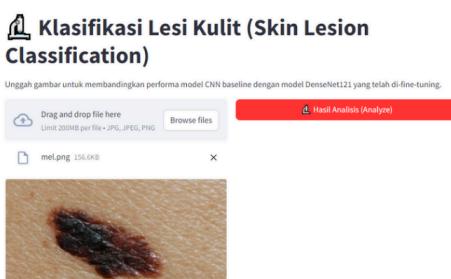
- Lapisan *GlobalAveragePooling2D* untuk mengurangi dimensi fitur gambar menjadi 1 vektor.
- Lapisan *Dense* yang merupakan lapisan terhubung dengan 128 lapisan dan aktivasi ReLU
- Lapisan *Dropout* untuk mencegah *overfitting* data dengan secara acak menghilangkan 50% neuron selama proses latihan.
- 7 lapisan keluaran yang *dense* dengan aktivasi *softmax*.

Menggunakan optimisasi Adam dengan *learning rate* lebih kecil yaitu 0.0001, dan *fine tuning* dengan mencairkan atau *unfreeze* 10 lapisan terakhir dari model DenseNet121 agar model yang digunakan dapat beradaptasi dengan dataset lesi kulit, lapisan ini akan dilatih bersamaan dengan lapisan yang ada, meningkatkan kecepatan dan performa akurasi dan *loss* dari hasil *training* model.

#### F. GUI Development

GUI dibuat dan dikembangkan untuk membuat proses klasifikasi jenis lesi kulit lebih mudah, efisien, interaktif, dan *user-friendly*. Pengguna dapat menggugah gambar lesi kulit yang kemudian menghasilkan keluaran prediksi jenis lesi kulit nya dari model *baseline* dan model *advanced results*.

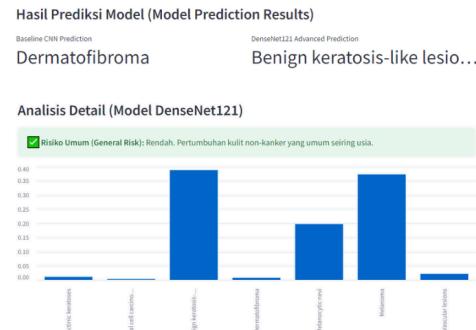
Gambar 5 menunjukkan *front-end framework* menggunakan Streamlit sebagai tampilan dari GUI nya, yang terdiri atas berbagai keterangan dan judul, tombol unggah, dan tombol untuk mengeluarkan hasil. Proses *back-end framework* menggunakan TensorFlow/Keras untuk membangun tampilan bagi pengguna. *Platform deployment* menggunakan *Ngrok Tunnel*, menggunakan token otorisasi untuk membuat URL publik sehingga pengguna dapat mengakses GUI menggunakan *link* tersebut.



Gambar 5. *Front-End Framework GUI*

Hasil GUI akan memberikan prediksi dari tiap model yang digunakan, yaitu CNN dan DenseNet121. Hasil dari model DenseNet121 dilengkapi dengan konteks medis mengenai deskripsi lesi kulit dan risiko, tingkat *alert* apakah sukses, *warning*, atau error, serta plot untuk probabilitas tiap prediksi pada 7 jenis lesi kulit. Pada gambar 6, setelah menggugah gambar

kedua model mungkin menunjukkan hasil yang berbeda, gambar 6 juga menunjukkan persentase probabilitas dari tiap lesi kulit ketika model DenseNet121 ingin memprediksi jenisnya.



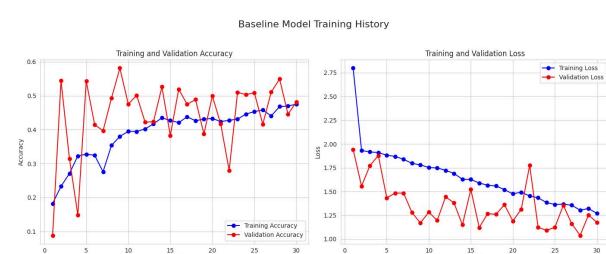
Gambar 6. Output Prediksi Dari Model

### III. HASIL

Kinerja dari kedua model yang digunakan, *baseline* Convolutional Neural Network (CNN) dan DenseNet121 yang telah di-*fine-tuning* dievaluasi menggunakan *unseen test set* yang terdiri dari 1503 gambar. Evaluasi ini difokuskan pada akurasi keseluruhan serta metrik *precision*, *recall*, dan *F1-score* untuk setiap kelas agar dapat menilai efektivitas model dalam skenario klasifikasi *multi-class* dengan data yang tidak seimbang (*imbalance*)

#### A. Kinerja Model CNN *Baseline*

Model *baseline* yang dibuat dari nol dan dilatih dengan *class weights* menunjukkan performa yang terbatas. Seperti yang terlihat pada Gambar. 7, kurva *training* dan *validation* untuk model ini tidak menunjukkan konvergensi yang baik. Akurasi validasi sangat tidak stabil dan terdapat celah yang besar dengan akurasi *training*, yang merupakan indikator klasik dari *overfitting* [6].

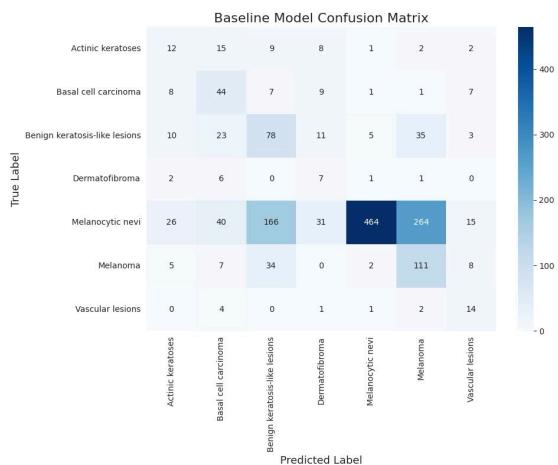


Gambar 7. Kurva akurasi dan *loss* *training* dan *validation* Model CNN *Baseline*

Pada *test set*, model ini hanya mencapai akurasi keseluruhan sebesar 49% dengan *macro average F1-score* 0.36. Tabel I menunjukkan bahwa meskipun *recall* untuk kelas **Melanoma** cukup tinggi (0.66), *precision*-nya sangat rendah (0.27), yang berarti model ini menghasilkan banyak sekali alarm palsu. *Confusion matrix* pada Gambar. 8 mengkonfirmasi hal ini, di mana terlihat adanya kebingungan yang signifikan antara kelas **Melanocytic nevi** dan **Melanoma**.

Tabel 1. Metrik evaluasi dari Model CNN Baseline

| Tipe Lesi            | Precision   | Recall      | F1-Score    | Jumlah Sampel |
|----------------------|-------------|-------------|-------------|---------------|
| Actinic keratoses    | 0.19        | 0.24        | 0.21        | 49            |
| Basal cell carcinoma | 0.32        | 0.57        | 0.41        | 77            |
| Benign keratosis     | 0.27        | 0.47        | 0.34        | 165           |
| Dermatofibroma       | 0.10        | 0.41        | 0.17        | 17            |
| Melanocytic nevi     | 0.98        | 0.46        | 0.63        | 1006          |
| <b>Melanoma</b>      | <b>0.27</b> | <b>0.66</b> | <b>0.38</b> | <b>167</b>    |
| Vascular lesions     | 0.29        | 0.64        | 0.39        | 22            |
| <b>Akurasi</b>       |             |             | <b>0.49</b> | <b>1503</b>   |
| <b>Macro Avg</b>     | <b>0.34</b> | <b>0.49</b> | <b>0.36</b> | <b>1503</b>   |
| <b>Weighted Avg</b>  | <b>0.74</b> | <b>0.49</b> | <b>0.53</b> | <b>1503</b>   |



Gambar 8. Confusion Matrix dari Model CNN Baseline

#### B. Kinerja Model DenseNet121 yang Telah di-Fine-Tuning

Model DenseNet121 yang menggunakan pendekatan *transfer learning* dan *fine-tuning* menunjukkan peningkatan performa yang signifikan. Berbeda dengan model *baseline*, hasil *training* untuk model DenseNet121 (Gambar 9) menunjukkan proses pembelajaran yang jauh lebih baik dan stabil. Kurva akurasi *training* dan validasi sama-sama meningkat secara konsisten, dan celah di antara keduanya jauh

lebih kecil. Hal ini menandakan bahwa model berhasil menggeneralisasi fitur dengan baik dan tidak mengalami *overfitting* yang parah. Arsitektur DenseNet yang efisien dalam *feature reuse* [7] terbukti sangat cocok untuk tugas ini.



Gambar 9. Kurva akurasi dan loss training dan validation Model CNN Baseline

Peningkatan performa terlihat jelas pada hasil evaluasi di *test set* (Tabel II dan Gambar. 10). Akurasi keseluruhan meningkat menjadi 86%, dan *macro average F1-score* meningkat lebih dari dua kali lipat menjadi 0.77. Ini menunjukkan bahwa model tidak hanya akurat, tetapi juga memiliki kinerja yang seimbang di semua tujuh kelas lesi, sebuah syarat penting untuk alat bantu diagnostik yang dapat diyakinkan [8].

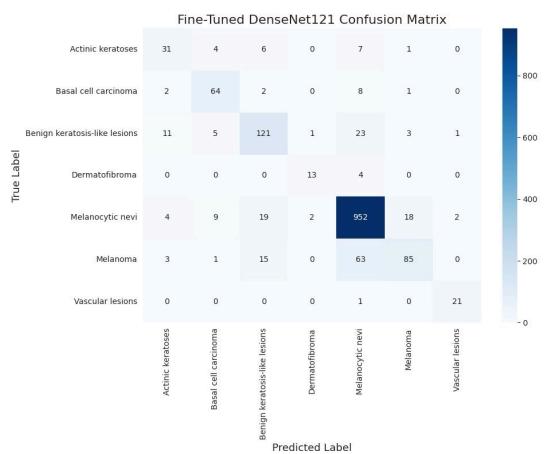
Untuk kelas Melanoma, seperti yang terlihat pada Tabel II, model ini mencapai *F1-score* 0.62. Yang paling menonjol, *precision*-nya meningkat hampir tiga kali lipat menjadi 0.79, dengan *recall* sebesar 0.51. Hal ini menunjukkan pergeseran perilaku model yang positif. *Precision* sebesar 0.79 berarti bahwa ketika model mengidentifikasi sebuah lesi sebagai potensi melanoma, prediksinya benar dalam 79% kasus, yang secara drastis mengurangi tingkat *false alarm* dibandingkan dengan model *baseline*. Meskipun *recall* 0.51 menunjukkan bahwa model ini masih melewatkannya sekitar setengah dari kasus melanoma yang sebenarnya, *precision*-nya yang tinggi menjadikannya alat bantu diagnostik yang jauh lebih dapat dipercaya bagi para klinisi.

*Confusion matrix* pada Gambar 10 mengilustrasikan peningkatan performa ini. Nilai di sepanjang diagonal utama jauh lebih kuat untuk semua kelas, menandakan jumlah klasifikasi yang benar lebih tinggi. Jumlah kasus Melanocytic nevi yang salah diklasifikasikan sebagai Melanoma turun dari 264 pada model *baseline* menjadi hanya 63, sementara jumlah kasus melanoma yang teridentifikasi dengan benar (*True Positives*) adalah 85.

Tabel II. Metrik evaluasi dari Model DenseNet121

| Tipe Lesi            | Precision | Recall | F1-Score | Jumlah Sampel |
|----------------------|-----------|--------|----------|---------------|
| Actinic keratoses    | 0.61      | 0.63   | 0.62     | 49            |
| Basal cell carcinoma | 0.77      | 0.83   | 0.80     | 77            |

|                     |             |             |             |             |
|---------------------|-------------|-------------|-------------|-------------|
| Benign keratosis    | 0.74        | 0.73        | 0.74        | 165         |
| Dermatofibroma      | 0.81        | 0.76        | 0.79        | 17          |
| Melanocytic nevi    | 0.90        | 0.95        | 0.92        | 1006        |
| <b>Melanoma</b>     | <b>0.79</b> | <b>0.51</b> | <b>0.62</b> | <b>167</b>  |
| Vascular lesions    | 0.88        | 0.95        | 0.91        | 22          |
| <b>Akurasi</b>      |             |             | <b>0.86</b> | <b>1503</b> |
| <b>Macro Avg</b>    | <b>0.79</b> | <b>0.77</b> | <b>0.77</b> | <b>1503</b> |
| <b>Weighted Avg</b> | <b>0.85</b> | <b>0.86</b> | <b>0.85</b> | <b>1503</b> |



Gambar x. *Confusion Matrix* dari Model DenseNet121

#### IV. DISKUSI

##### A. Analisis Perbandingan Performa Model

Model DenseNet121 yang diimplementasikan dengan strategi transfer learning dan fine-tuning mencapai akurasi sebesar 86% jauh melampaui model CNN baseline yang hanya mencapai akurasi 49%. Keunggulan ini disebabkan oleh cara kerja DenseNet121. Konsep transfer learning dimana model baseline harus mempelajari fitur visual dari nol, mulai dari deteksi tepi, tekstur, hingga bentuk-bentuk kompleks, hanya dengan berbekal dataset HAM10000. Sebaliknya, model DenseNet121 memulai dengan pemahaman awal yang sangat luas karena telah dilatih pada jutaan gambar dari dataset ImageNet [9].

Meskipun dataset ImageNet tidak mengandung gambar lesi kulit sama sekali, kemampuannya terletak pada pengetahuan fundamental yang diajarkannya. Saat dilatih pada jutaan gambar objek sehari-hari, sebuah model tidak hanya belajar mengenali 'kucing' atau 'mobil', tetapi ia membangun pemahaman visual secara bertingkat. Pada lapisan-lapisan awalnya, ia menjadi ahli dalam mendeteksi fitur-fitur universal tingkat rendah seperti garis, tepian, dan gradien warna. Pada lapisan tengah, ia belajar menggabungkan

fitur-fitur dasar ini untuk mengenali pola yang lebih kompleks seperti berbagai macam tekstur. Konsep *transfer learning* bekerja dengan 'meminjam' keahlian deteksi fitur universal ini. Kita membuang lapisan akhir model yang bertugas mengenali objek spesifik, namun mempertahankan seluruh lapisan dibawahnya yang sudah sangat mahir dalam 'melihat' tekstur, bentuk, dan tepian. Dengan demikian, kita tidak mentransfer pengetahuan tentang *apa* yang dilihat model (objek), melainkan mentransfer *cara* ia melihat (kemampuan deteksi fitur), lalu mengajari kembali kemampuan tersebut untuk diaplikasikan pada konteks baru yang sangat spesifik, yaitu klasifikasi lesi kulit [10].

Selain itu DenseNet121 memiliki konsep feature reuse dimana setiap lapisan dalam DenseNet menerima input dari semua lapisan sebelumnya, memungkinkan penggunaan kembali fitur secara ekstensif. Untuk citra medis di mana detail tekstur dan pola yang halus, kemampuan untuk terus-menerus mengakses dan menggabungkan fitur level rendah dan tinggi ini memungkinkan model membangun representasi yang lebih kaya dan informatif sehingga klasifikasi yang jauh lebih unggul dibandingkan model CNN baseline [11].

##### B. Analisis Tantangan pada Kelas Minoritas

Analisis pada classification report dan confusion matrix menunjukkan adanya tantangan persisten dalam mendeteksi beberapa kelas, terutama Melanoma. Model ini hanya mencapai recall sebesar 0.51 untuk kelas Melanoma, yang berarti hampir separuh kasus Melanoma pada data uji gagal terdeteksi dan salah diklasifikasikan, seringkali sebagai Melanocytic nevi. Kegagalan ini dapat dijelaskan oleh kombinasi dua faktor utama. Faktor pertama adalah kemiripan visual yang tinggi. Melanoma pada tahap awal seringkali memiliki karakteristik visual yang sangat mirip dengan Melanocytic nevi. Ketika fitur-fitur pembeda seperti asimetri atau variasi warna sangat halus, model bahkan yang canggih sekalipun dapat kesulitan untuk membuat keputusan yang tepat [12].

Faktor kedua adalah masalah ketidakseimbangan data yang ekstrim. Walaupun teknik seperti class\_weight telah diterapkan untuk memberikan penalti lebih besar saat salah mengklasifikasikan kelas minoritas, hal ini tidak sepenuhnya menghilangkan bias. Model telah melihat ribuan contoh Melanocytic nevi yang membentuk "jalur prediksi default" yang sangat kuat. Ketika dihadapkan pada gambar yang ambigu, meskipun diberi peringatan melalui class\_weight, model mungkin masih cenderung pada prediksi kelas mayoritas yang paling familiar baginya. Dengan demikian, kesulitan dalam mendeteksi Melanoma secara akurat merupakan hasil dari tumpang tindih fitur visual dengan kelas mayoritas dan bias statistik yang persisten akibat ketidakseimbangan data.

##### C. Limitations

Meskipun model DenseNet121 menunjukkan peningkatan performa yang signifikan, penting untuk mengakui beberapa batasan yang ada dalam proyek ini. Keterbatasan utama terletak pada dataset yang digunakan, yang memiliki ketidakseimbangan kelas yang signifikan dan kurangnya diversitas pada warna kulit pasien. Selain itu, model yang dikembangkan berfungsi sebagai black box yang mampu memberikan prediksi akurat namun tidak dapat menjelaskan alasan klinis di baliknya. Proyek ini juga tidak mempertimbangkan riwayat medis pasien sebagai salah satu faktor masukan. Oleh karena itu, perlu ditekankan bahwa sistem ini dirancang murni sebagai alat bantu pendukung keputusan dan bukan merupakan pengganti diagnosis profesional yang dilakukan oleh dokter yang berkualifikasi.

#### D. Take-home Message

Pemilihan arsitektur model yang canggih saja tidak cukup untuk menjamin keberhasilan. Keberhasilan tugas klasifikasi ini juga sangat bergantung pada kemampuan untuk menangani tantangan data secara tepat, yaitu melalui teknik seperti augmentasi data dan penanganan ketidakseimbangan kelas. Pertimbangan etikal dalam pengembangan sistem AI juga penting. Memahami dan mengkomunikasikan batasan sistem secara transparan merupakan kunci untuk membangun kepercayaan dan memastikan pemanfaatan teknologi AI yang bertanggung jawab di bidang medis.

#### E. Future Work

Untuk pengembangan di masa depan, beberapa langkah dapat diambil untuk meningkatkan dan memvalidasi sistem ini. Pertama, eksplorasi arsitektur State-of-the-Art lainnya dapat dilakukan sebagai opsi pembanding model yang lebih lanjut. Kedua, model dapat ditingkatkan dengan menginkorporasikan metadata pasien terstruktur seperti usia, jenis kelamin, dan lokasi lesi sebagai input tambahan untuk memperkaya konteks prediksi. Terakhir, langkah validasi yang paling krusial adalah dengan menguji model yang telah dilatih pada beberapa dataset eksternal. Dataset ini idealnya mencakup rentang warna kulit dan kondisi akuisisi gambar yang lebih beragam untuk memastikan ketangguhan (robustness) dan keadilan (fairness) model sebelum dipertimbangkan untuk aplikasi dunia nyata.

## V. KESIMPULAN

Proyek ini berhasil mengembangkan model ML untuk klasifikasi 7 kelas lesi kulit untuk deteksi dini kanker kulit. Menggunakan dataset HAM10000, proyek ini membandingkan performa antara model CNN yang sederhana dan DenseNet121 yang lebih canggih melalui pendekatan *transfer learning* dan *fine-tuning*.

Hasil evaluasi pada data uji menunjukkan bahwa model DenseNet121 yang di-*fine-tuning* berhasil mencapai akurasi

86%, jauh lebih tinggi dibandingkan dengan model CNN baseline yang hanya mencapai 49%. Penerapan teknik *data augmentation* dan *class weighting* berhasil menangani ketidakseimbangan data dalam dataset, meningkatkan performa model.

Secara keseluruhan, proyek ini menunjukkan potensi besar penggunaan ML. Sistem yang dikembangkan dapat meningkatkan kecepatan diagnosis, mengurangi kesalahan diagnostik, dan memperluas akses terhadap skrining kanker kulit yang andal, terutama di wilayah yang kekurangan tenaga spesialis.

Namun, meskipun sistem ini menunjukkan performa yang baik, masih ada tantangan dalam mendeteksi beberapa kelas lesi kulit, terutama Melanoma, yang memerlukan perbaikan lebih lanjut. Selain itu, sistem ini hanya berfungsi sebagai alat bantu keputusan dan bukan pengganti diagnosis medis profesional.

Proyek ini juga memiliki beberapa keterbatasan, seperti ketidakseimbangan data yang ekstrem dan kurangnya diversitas warna kulit, yang perlu diperhatikan dalam pengembangan lebih lanjut. Kedepannya, dapat dilakukan eksplorasi arsitektur model lainnya serta integrasi metadata pasien seperti usia dan jenis kelamin yang dapat memperkaya konteks prediksi dan meningkatkan akurasi model.

## DAFTAR ACUAN

- [1] Sathe NC, Zito PM. Skin Cancer. [Updated 2025 Feb 17]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK441949/>.
- [2] GeeksforGeeks, “What is Exploratory Data Analysis?,” GeeksforGeeks, May 10, 2025. <https://www.geeksforgeeks.org/r-data-analysis/what-is-exploratory-data-analysis/> (accessed Jun. 13, 2025).
- [3] “Skin Cancer MNIST: HAM10000,” Kaggle, Sep. 20, 2018. [https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000/data?select=HAM10000\\_metadata.csv](https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000/data?select=HAM10000_metadata.csv) (accessed Jun. 13, 2025).
- [4] 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, Jan. 2023, doi: 10.5336/biostatic.2022-93961.
- [5] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” Scientific Data, vol. 5, no. 1, Aug. 2018, doi: 10.1038/sdata.2018.161.
- [6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014, Available: <https://jmlr.org/papers/v15/srivastava14a.html>
- [7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Jul. 2017, doi: <https://doi.org/10.1109/cvpr.2017.243>.
- [8] H. Deng, H. Qiao, Q. Dai, and C. Ma, “Deep learning in photoacoustic imaging: a review,” *Journal of Biomedical Optics*, vol. 26, no. 04, Apr. 2021, doi: <https://doi.org/10.1117/1.jbo.26.4.040901>.

- [9] Iamtapendu, "Introduction to DenseNet-121," Kaggle, May 11, 2025. Available: <https://www.kaggle.com/code/iamtapendu/introduction-to-densenet-121>
- [10] Z. Aslan, "Transfer Learning using DenseNet201 - Zeynep Aslan - Medium," *Medium*, Mar. 16, 2022. [Online]. Available: <https://medium.com/@zeynepaslann/transfer-learning-using-densenet201-525749762ca9>
- [11] B. C, "How DenseNet reuses features to build smarter networks," *Medium*, Jun. 11, 2025. [Online]. Available: [https://medium.com/@caring\\_smitten\\_gerbil\\_914/how-densenet-reuses-features-to-build-smarter-networks-1294f3862498](https://medium.com/@caring_smitten_gerbil_914/how-densenet-reuses-features-to-build-smarter-networks-1294f3862498)
- [12] P. Fräntti and R. Marinescu-Istodor, "Soft precision and recall," *Pattern Recognition Letters*, vol. 167, pp. 115–121, Feb. 2023, doi: 10.1016/j.patrec.2023.02.005.

#### LAMPIRAN

##### A. Ketersediaan Kode

Seluruh kode sumber untuk proyek ini, termasuk *notebook* untuk *training* dan evaluasi model serta skrip untuk *deployment* GUI, tersedia di repositori GitHub publik pada tautan berikut:

[https://github.com/HasanLathif/Final-Project-PSB-Gro up10](https://github.com/HasanLathif/Final-Project-PSB-Group10)