

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker kulit merupakan penyakit yang disebabkan oleh paparan radiasi *ultraviolet* (UV), terutama dari sinar matahari. Paparan sinar UV menyebabkan kerusakan DNA, mutasi gen, dan melemahkan sistem kekebalan kulit (Narayanan dkk., 2010). Kanker kulit adalah salah satu jenis kanker paling umum dan berbahaya di dunia, terutama jika tidak terdeteksi sejak dini dan diobati secara tepat. Kanker ini termasuk dalam kategori kanker yang dapat menyebabkan tingkat kematian tinggi, terutama pada stadium lanjut (Sung dkk., 2021). Faktor utama yang menyebabkan meningkatnya prevalensi kanker kulit adalah paparan sinar UV yang terus menerus, terutama bagi individu dengan kulit terang. Risiko ini diperburuk oleh faktor genetis dan lingkungan, seperti riwayat keluarga atau mutasi pada gen tertentu yang mempengaruhi regulasi siklus sel kulit (Hossain dkk., 2024).

Secara umum, diagnosis kanker kulit dilakukan melalui prosedur biopsi dan analisis mikroskopis. Prosedur biopsi melibatkan pengambilan sampel jaringan kecil dari area yang dicurigai mengandung sel-sel kanker untuk dianalisis secara mendalam oleh dokter spesialis atau ahli dermatopatologi. Teknik diagnostik ini memerlukan waktu yang signifikan untuk memperoleh hasil yang akurat, serta membawa risiko potensial terjadinya komplikasi atau cedera selama proses pengambilan sampel biopsi (Esteva dkk., 2017).

Penelitian dalam pengembangan model prediksi dini kanker kulit menggunakan kombinasi teknik *deep learning* dan analisis citra dermoskopi memiliki potensi untuk memberikan solusi objektif dan konsisten bagi para dokter dalam mengidentifikasi kanker kulit secara lebih akurat (Esteva dkk., 2017). *Deep learning* digunakan secara efektif dalam mengklasifikasikan penyakit melalui citra medis termasuk deteksi kanker, analisis retina, dan segmentasi jaringan otak. *Deep learning* berkerja dengan mengotomatiskan ekstraksi fitur dari citra medis dan mendeteksi pola-pola yang relevan. *Deep learning* dalam kapasitasnya untuk belajar dari data mentah, membuka peluang bagi aplikasi medis yang lebih efisien dan akurat (LeCun dkk., 2015) salah satunya implementasi dalam deteksi kanker kulit.

Penelitian terdahulu telah dilakukan untuk mengkaji klasifikasi penyakit kanker kulit. Telah menerapkan berbagai metode dalam upaya prediksi klasifikasi jenis kanker kulit yang dialami pasien menggunakan *dataset* HAM10000 (Tschandl dkk., 2018). Metode-metode tersebut melibatkan penggunaan teknologi canggih seperti pengolahan citra menggunakan arsitektur *deep learning* terbaru maupun algoritma terbaru. Sebagai contoh penelitian yang dilakukan oleh Smith dkk yaitu dengan arsitektur *inception-V3* dan mendapatkan akurasi 81% (Smith dkk., 2024). Terdapat pendapat arsitektur *deep learning* yang memiliki tingkat akurasi lebih tinggi menggunakan arsitektur yang lebih baru dibandingkan dengan model sebelumnya. hal ini dinyatakan oleh LeCun (2015) yang menyatakan bahwa yaitu arsitektur *deep learning* yang lebih baru memungkinkan pembelajaran representasi data dalam beberapa tingkat abstraksi, menangkap fitur kompleks dari data mentah. Serta arsitektur modern dengan lebih banyak lapisan (*depth*) memungkinkan jaringan neural mempelajari fungsi kompleks. Model lebih dalam, seperti *ResNet*, *EfficientNet*, dan *VGG*, terbukti lebih efisien dalam memproses data kompleks dan menghasilkan hasil yang lebih baik pada klasifikasi gambar dan pemrosesan bahasa alami. Berikutnya dilakukan oleh Moturi dkk model menggunakan *MobilenetV2* dan mencapai akurasi 85% (Moturi dkk., 2024). Hasan dkk. (2024) juga melakukan penelitian mengenai klasifikasi kanker kulit dan mencapai akurasi 91% dengan menggunakan arsitektur gabungan. Selanjutnya oleh Thwin dkk. (2024) menggunakan model *ensemble* pada *dataset* yang sama mencapai akurasi 90%.

EfficientNetV2 secara signifikan meningkatkan akurasi sekaligus menjaga biaya komputasi tetap rendah. Model ini menggabungkan pencarian arsitektur jaringan saraf yang dioptimalkan untuk pelatihan dan pendekatan pembelajaran progresif. Berdasarkan penelitian Sunil dkk. (2022) Model ini mencapai akurasi deteksi sebesar 98.26% pada *dataset* daun kapulaga, menunjukkan peningkatan akurasi yang signifikan dibandingkan model CNN konvensional dan *EfficientNet* biasa. Pada penelitian (Devi dkk., 2023) *EfficientNetV2* mengungguli model lain seperti *InceptionV3* dan *Vision Transformer*, dengan akurasi mencapai 99.5% pada *dataset cassava*, 97.5% pada *PlantVillage*, dan 80.1% pada *dataset IP102*, dengan waktu pelatihan yang lebih singkat. Kedua penelitian ini menegaskan keunggulan *EfficientNetV2* dalam memaksimalkan akurasi klasifikasi dengan efisiensi komputasi yang tinggi, menjadikannya lebih efektif dalam proses *deep learning*. Kecepatan pelatihan menjadi lebih cepat dan ukuran parameter berkurang hingga 6,8 kali lipat.

Progressive learning menyesuaikan ukuran gambar dan regulasi selama pelatihan, sehingga model lebih efisien. Hasilnya, *EfficientNetV2* mampu mengoptimalkan waktu pelatihan dan efisiensi parameter tanpa mengorbankan akurasi (Tan & Le, 2021).

Klasifikasi hierarkikal adalah metode klasifikasi multi kelas yang dimana kelas target dari klasifikasi disusun dalam sebuah hierarki umumnya sebagai sebuah *tree* atau *directed acyclic graph* dan klasifikasi dilakukan berdasarkan struktur yang sudah didefinisikan sebelumnya. Klasifikasi hierarkikal memanfaatkan struktur kelas untuk menangkap hubungan antar kelas, mencegah kesalahan klasifikasi dengan menggunakan informasi lokal lebih efektif menyempurnakan prediksi dalam setiap *step* dalam hierarki. Dalam klasifikasi *flat*, model sering kesulitan dengan banyak kelas karena harus membedakan semua kategori tanpa panduan struktur hierarki, yang dapat meningkatkan kesalahan klasifikasi, terutama pada *dataset* yang kompleks. Sebaliknya, klasifikasi hierarkis mengatasi masalah ini dengan memanfaatkan hubungan antar kelas, sehingga menghasilkan generalisasi yang lebih baik dan prediksi yang lebih akurat (Silla & Freitas, 2011).

Pada penelitian Pereira dkk. (2020) model hierarkis digunakan untuk mengelompokkan pneumonia berdasarkan jenis patogen (virus, bakteri, jamur), kemudian lebih lanjut membagi kelompok virus menjadi subkategori seperti COVID-19, SARS, dan MERS. Klasifikasi ini mengikuti struktur hierarki berjenjang, di mana kelas pneumonia diatur berdasarkan urutan organisme penyebabnya. Penerapan klasifikasi hierarkis dalam studi ini terbukti efektif, dengan model mencapai skor F1 sebesar 0.89 dalam mengenali kasus COVID-19. Keberhasilan ini menyoroti keunggulan pendekatan hierarkis dalam memperbaiki akurasi deteksi untuk data yang memiliki ketidakseimbangan kelas. Penelitian Suleiman dkk. (2024) menggunakan model klasifikasi biner hierarkis dua langkah dengan memanfaatkan *transfer learning* dan algoritma *random forest* untuk mengatasi ketidakseimbangan kelas dalam *dataset* lesi kulit. Pada tahap pertama, model memisahkan lesi menjadi "benign" dan "lainnya," sedangkan tahap kedua mengklasifikasikan lebih lanjut "lainnya" menjadi melanoma atau seborrheic keratosis. Dengan pendekatan ini, kombinasi *DenseNet121* sebagai ekstraktor fitur dan *random forest* sebagai pengklasifikasi mencapai akurasi *multiclass* seimbang (BMA) sebesar 91,07%, lebih tinggi dibandingkan model *deep learning* murni, membuktikan efektivitas pendekatan hierarkis dalam klasifikasi lesi kulit.

Transfer learning adalah pendekatan dalam pembelajaran mesin yang bertujuan untuk meningkatkan performa model pada domain target dengan memanfaatkan informasi yang telah diperoleh dari domain sumber yang terkait. Teknik ini berguna ketika data pelatihan untuk domain target terbatas, mahal untuk diperoleh, atau sulit diakses. *Transfer learning* mendukung *robustness* model dengan cara mengintegrasikan pola-pola yang telah dipelajari sebelumnya dari domain sumber sehingga model dapat menangani domain target dengan lebih baik, meskipun terdapat perbedaan distribusi data antara keduanya (Weiss dkk., 2016).

Berdasarkan studi yang telah ditinjau, metode klasifikasi hierarkikal belum diterapkan dalam kasus klasifikasi kanker kulit ini. Dilakukannya penelitian ini agar dapat melihat bagaimana akurasi berhubungan dengan metode hierarkikal didasari untuk “menyeimbangkan” data tanpa harus membuang dari kelebihan data yang tersedia (Vens dkk, 2008). Arsitektur *EfficientNetV2* dipilih karena dirancang untuk meringankan beban komputasi melalui blok yang disebut *Fused MBConv*. Dalam kasus klasifikasi hierarkis yang membutuhkan model lebih dari satu, penggunaan *EfficientNetV2* menjadi pilihan yang tepat. Serta untuk meningkatkan kemampuan *robustness* setiap model juga menerapkan *transfer learning* (Tan & Le, 2021).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dituliskan, permasalahan yang ada didalam penelitian ini adalah:

1. Bagaimana melakukan klasifikasi hierarkikal penyakit kanker kulit melalui data citra kanker kulit menggunakan metode *hierarchical classification* menggunakan arsitektur *EfficientNetV2B0* dengan *transfer learning* ?
2. Bagaimana akurasi model untuk klasifikasi kanker kulit yang didapatkan dengan menggunakan metode *hierarchical classification* menggunakan arsitektur *EfficientNetV2B0* dengan *transfer learning* ?

1.3 Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah dituliskan, tujuan yang ingin dicapai pada penelitian ini adalah :

1. Mengimplementasikan klasifikasi hierarkikal penyakit kanker kulit melalui data citra kanker kulit dengan menggunakan metode *hierarchical classification* menggunakan arsitektur *EfficientNetV2B0* dengan *transfer learning*.
2. Mengevaluasi akurasi model yang didapatkan dengan menggunakan metode *hierarchical classification* dan metode *transfer learning* menggunakan arsitektur *EfficientNetV2B0* dengan *transfer learning*.

Metode klasifikasi hierarkikal yang menggunakan arsitektur *EfficientNetV2B0* memiliki potensi besar dalam meningkatkan akurasi identifikasi kanker kulit pada citra medis. Dengan memanfaatkan model ini, hasil klasifikasi menjadi lebih presisi, yang penting dalam mendeteksi jenis kanker kulit sejak dini. Penelitian ini diharapkan dapat memberikan kesempatan bagi penderita kanker kulit untuk mendapatkan diagnosis yang lebih cepat dan akurat. Diagnosis yang lebih awal dapat meningkatkan peluang pasien menerima perawatan yang tepat waktu, sehingga dapat mengurangi risiko komplikasi atau kerusakan yang lebih serius pada tahap lanjut.

1.4 Ruang Lingkup

Agar tidak menyimpang dari permasalahan, maka dari itu penelitian ini memiliki ruang lingkup penelitian :

1. *Dataset* yang digunakan merupakan *dataset* sekunder, diperoleh dari laman *kaggle* dengan nama *dataset* “*Skin Cancer MNIST: HAM10000*”. Data ini merupakan data yang telah dikurasi dari *Harvard Dataverse*, yang dimana *dataset* dibuat oleh *Medical University of Vienna* yang terdiri dari 10.015 citra.
2. Jumlah data yang digunakan yaitu 10.015 yang terbagi menjadi 7 kelas yaitu *Melanocytic nevi (nv)*, *Melanoma (mel)*, *Benign keratosis (bkl)*, *Basal cell carcinoma (bcc)*, *Actinic keratoses (akiec)*, *Vascular (vasc)*, *Dermatofibroma (df)*. Kemudian data dibagi menjadi tiga bagian yaitu *training*, *validation*, dan *testing*.
3. Tahap *preprocessing* data, pelatihan, dan pengujian dilakukan di lingkungan *jupyter notebook* pada platform *google colab*.
4. Penerapan klasifikasi hierarkikal dan *transfer learning* menggunakan *library TensorFlow*
5. *Hyperparameter* yang dicari kombinasi terbaiknya meliputi *learning rate*, *optimizer*, dan *data augmentation* untuk mendapatkan akurasi *testing* yang paling besar.

1.5 Sistematika Penulisan

Sistematika penulisan menjadi pada penelitian ini terbagi menjadi lima bab dengan uraian :

BAB I. PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan laporan tugas akhir mengenai klasifikasi hierarkikal.

BAB II. TINJAUAN PUSTAKA

Bab ini membahas teori yang digunakan untuk melakukan penelitian ini. Pada bab ini membahas beberapa yaitu Studi literatur, Citra Digital, Klasifikasi, *Transfer Learning*, *Augmentasi Data*, *Convolution Neural Network*, *Arsitektur EfficientNetV2*, *Loss Function*, *Backpropagation*, *Adam*, *SGD*, *Dense* dan *Confusion Matrix*.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai metodologi penelitian yang digunakan dalam menentukan nilai *hyperparameter*, pembagian tingkat hierarki untuk melakukan klasifikasi penyakit kanker kulit dengan penerapan klasifikasi hierarkikal menggunakan arsitektur *EfficientNetV2*. Proses yang dilakukan dalam penelitian ini adalah pengumpulan data, *preprocessing data*, penentuan nilai *hyperparameter*, pembangunan model, pelatihan model, dan evaluasi model.

BAB IV. HASIL DAN PEMBAHASAN

Bab ini membahas tentang lingkungan dan perangkat yang digunakan dalam penelitian dan memberikan analisa hasil pelatihan dan pengujian dari model terbaik.

BAB V. PENUTUP

Bab ini berisi kesimpulan dari bab-bab yang telah dibahas sebelumnya dan saran untuk penelitian selanjutnya mengenai klasifikasi kanker kulit.