

Kombinasi U-Net DenseNet untuk Klasifikasi Dermoskopi Kanker Kulit

Muhammad Nadine Zuhdan Azra
Fakultas Informatika, Universitas Telkom, Bandung
nadinezuhdan@student.telkomuniversity.ac.id

Tjokorda Agung Budi Wirayuda
Fakultas Informatika, Universitas Telkom, Bandung
cokagung@telkomuniversity.ac.id

Kanker kulit merupakan jenis kanker yang paling umum ditemukan. Namun kanker kulit masih dapat merusak struktur organ hingga menyebabkan kematian. Resiko kanker kulit dapat dimitigasi jika dapat ditangani sejak awal. Oleh karena itu mendeteksi apabila seseorang terkena melanoma sangatlah penting. Salah satu metode yang umum digunakan ahli dermatologi untuk mendeteksi melanoma adalah dengan melakukan dermoskopi dan mengujinya dengan bantuan algoritma visi komputer. Pada penelitian ini dibangun sebuah model kombinasi segmentasi dengan klasifikasi menggunakan U-Net dan DenseNet. Model ini bekerja dengan pertama melakukan segmentasi terhadap *input* menggunakan U-Net yang kemudian hasil dari model tersebut akan digunakan untuk melakukan *masking* pada *input* untuk model klasifikasi. Pelatihan model dilakukan menggunakan dataset HAM10000 dan dilakukan secara bertahap. Pelatihan dimulai dari model U-Net, kemudian langsung digabungkan dengan model kombinasi untuk mengevaluasi performa dari model serta metode *masking* yang digunakan. Pada akhirnya diperoleh sebuah model kombinasi dengan nilai akurasi 76%, dengan presisi rata-rata berbobot 76%, recall rata-rata berbobot 76%, dan F1-Score rata-rata berbobot 75% beserta model segmentasi dengan mean IoU sebesar 87%, F1-Score rata-rata sebesar 92%, Sensitivitas 94%, dan specificity 97%.

Kata kunci: kanker kulit, densenet, unet, segmentasi, klasifikasi, masking

Skin cancer is a common type of cancer. But skin cancer are still capable to ruin organ structures which may lead to death. The risk of skin cancer can be mitigated if a diagnosis was done early. Therefore an early melanoma diagnosis is very important. A common method dermatologist use to diagnose melanoma is to do a dermoscopy then testing it with the help of a computer vision algorithm. In this research a segmentation and classification combination model will be created using U-Net for segmentation and DenseNet as the classifier. The model works by segmenting the input using U-Net then using the mask on the input data before feeding it into the classifier. The model will be trained using the HAM10000 dataset and will be trained in steps. Starting with the U-Net model, before combining the trained model with the classifier to evaluate the model's performance alongside the masking method used. In the end a model with an accuracy of 76%, weighted average precision of 76%, weighted average recall of 76%, and a weighted average F1-Score of 75% alongside a segmentation model with a mean IoU of 87%, a mean average F1-Score of 92%, sensitivity of 94%, and specificity of 97%.

Keywords: skin lesion, densenet, unet, segmentation, classification, masking

1. Pendahuluan

Latar Belakang

Kanker kulit merupakan jenis kanker yang paling umum ditemukan di akhir abad ke-20, dan hingga sekarang masih pula meningkat jumlah penderita kanker jenis ini. Peningkatan jumlah pasien kanker kulit dapat diatribusikan kepada perubahan iklim yang meningkatkan jumlah radiasi ultraviolet yang diterima oleh seseorang [1]. Kanker kulit umumnya dibagi kedalam dua kategori, malignant melanoma, dan Non-melanoma Skin Cancer (NMSC). NMSC terdiri atas dua jenis kanker, yaitu Basal Cell Carcinoma (BCC) dan Squamous Cell Carcinoma (SCC), BCC merupakan jenis kanker kulit yang paling umum, mewakili sekitar 80% dari seluruh kasus kanker kulit NMSC, tipe kanker kulit ini dapat muncul tanpa kemunculan lesi terlebih dahulu [2]. Meskipun BCC sangat umum ditemukan, kanker kulit ini cenderung lebih jinak dibandingkan kanker kulit lain, namun BCC masih bersifat destruktif dan dapat merusak struktur sel sekitarnya [3]. SCC merupakan salah satu jenis kanker kulit NMSC, SCC pada awalnya muncul sebagai luka lesi biasa yang kemudian berkembang menjadi sebuah kanker. Dibandingkan dengan BCC, SCC bersifat lebih ganas, dengan kemungkinan metastasis sebesar 3-10%, sehingga kanker kulit ini dapat menyebar dan mengakibatkan resiko fatal. Terakhir terdapat jenis kanker Malignant melanoma, kanker kulit ini hanya mewakili 4% dari seluruh kasus kanker kulit, namun sekitar 65% dari seluruh kasus kematian akibat kanker kulit disebabkan oleh Malignant melanoma. Seperti SCC, Malignant melanoma dapat menyebar dengan cepat dan bersifat fatal apabila dibiarkan terlalu lama [2].

Untuk melakukan identifikasi kanker kulit, metode yang saat ini sedang dikembangkan adalah dengan melakukan dermatoskopi yang dibantu oleh sebuah algoritma computer vision [4]. Metode untuk melakukan diagnosa kanker kulit umumnya dibuat sembari mengikuti beberapa aturan dasar, seperti ABCD yang membantu untuk membedakan antara tipe kanker kulit malignant atau benign, dengan menggunakan pendekatan machine learning seperti Support Vector Machine (SVM) dan bahkan dengan menggabungkan teknik segmentasi menggunakan Convolutional Neural Network (CNN) diikuti oleh fitur ekstraksi dan terakhir melakukan klasifikasi oleh beberapa model classifier, namun dengan hasil yang kurang memuaskan [5]. Tetapi belakangan ini mengikuti kemunculan dari Deep Learning sebuah pendekatan machine learning yang telah

menunjukkan hasil jauh lebih baik dibandingkan dengan pendekatan tradisional [6]. Dengan menggunakan pendekatan Deep Learning, CNN dapat menghasilkan performa yang dapat menandingi ahli dermoskopi, dan jika dibandingkan dengan pendekatan sebelumnya, merupakan peningkatan yang sangat besar [5].

Untuk penelitian ini, model classifier dibangun berdasarkan paper Adegun, et al, yang memanfaatkan *DenseBlock* dari DenseNet Connected Convolutional Network (DenseNet) dengan pendekatan Fully Convolutional Network (FCN), model tersebut terbukti memiliki performa sangat baik saat digunakan dengan model segmentasi [5]. Kemudian segmentasi dilakukan dengan menggunakan model U-Net yang akan dioptimisasi kedalamannya, arsitektur tersebut memiliki kemampuan segmentasi yang baik, serta dapat melakukan ekstraksi fitur dengan baik [7].

Topik dan Batasannya

Penelitian ini berfokus kepada pencarian metode *masking* yang efektif untuk menggabungkan model segmentasi dengan model klasifikasi untuk klasifikasi hasil dermoskopi kanker kulit. Layanan Kaggle dan Google Colab dipakai untuk menjalankan penelitian dengan masing-masing memiliki spesifikasi GPU P100 VRAM 15 GB dan RAM 15 GB kemudian GPU L4 VRAM 16 GB dan RAM 15 GB, keduanya memiliki batas sesi 12 jam dengan Kaggle memiliki batas penggunaan GPU sebanyak 30 jam setiap pekan. Tuning parameter bagi model segmentasi dan klasifikasi akan dilakukan dalam batasan sistem yang digunakan. Dataset yang digunakan diperoleh melalui situs Kaggle berjudul Skin cancer: HAM10000. Dataset terdiri atas 10015 pasangan citra dermoskopi dan *mask* dengan panjang 600px dan tinggi 450px [8].

Tujuan

Penelitian ini bertujuan untuk menyusun sebuah model hasil kombinasi dari model segmentasi berupa U-Net dan klasifikasi dengan DenseNet yang memiliki performa baik. Tujuan tersebut dapat diuraikan lagi menjadi dua bagian yaitu, untuk menentukan konfigurasi terbaik masing-masing model, serta metode *masking* paling efektif untuk menghubungkan kedua model.

2. Studi Terkait

Adegun A et al. [5] membangun sistem dengan arsitektur DenseNet berbasis FCN untuk melakukan klasifikasi dengan bantuan segmentasi menggunakan arsitektur encoder-decoder dengan FCN sekaligus mengimplementasikan layer CRF untuk meningkatkan kualitas hasil segmentasi. Sistem ini diuji menggunakan dataset HAM10000 dan ISBI 2017 sebagai data training yang akan diaugmentasi, dengan dataset ISBI 2018 dan PH2 sebagai data test. Model terbaik diperoleh dengan nilai akurasi, F1-Score, dan Recall sebanyak 98%, 98%, dan 98.5%.

Anand V et al. [9] berhasil mengimplementasikan arsitektur U-Net untuk meningkatkan performa classifier dengan arsitektur CNN. Sistem ini menggunakan U-Net untuk melakukan segmentasi dan menggabungkan hasil segmentasi tersebut sebagai alpha layer untuk input model CNN. Dataset yang digunakan untuk menguji sistem ini adalah HAM10000, yang berisi 10015 citra dermoskopi yang memiliki 7 label berbeda. Sistem yang dibuat memperoleh hasil akurasi dengan nilai 97.96%, jauh lebih tinggi dari model-model yang tidak menggunakan segmentasi, meskipun hanya menggunakan arsitektur CNN biasa.

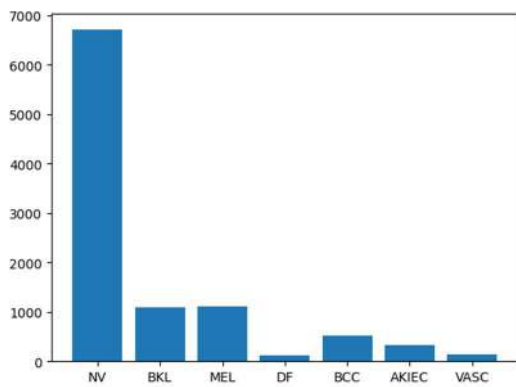
Fábio Perez et al [10] melakukan analisis terhadap berbagai metode augmentasi untuk dataset kanker kulit terhadap performa model klasifikasi. Eksperimen dilakukan dengan menggunakan 12 konfigurasi alur augmentasi, dari augmentasi dasar seperti *random flip* atau *random cropping* hingga gabungan dari 4 atau 5 augmentasi berturut-turut. Dari hasil eksperimen ditemukan alur augmentasi yang paling efektif dan konsisten merupakan alur yang berisi *random crops*, transformasi *affine*, *random flips*, dan augmentasi warna (perubahan saturasi, kontras, kecerahan, beserta *hue*), dengan metode affine, random flips, serta augmentasi warna memiliki performa yang konsisten apabila digunakan secara individu, dan *random crops*, *random erasing*, dan *lesion mix*, memiliki dampak yang mengurangi performa model apabila digunakan secara individu.

Dapat disimpulkan bahwa dengan melakukan optimasi terhadap model segmentasi, model klasifikasi dapat dibuat lebih efisien dan tetap mencapai performa yang mendekati model kompleks. Pada penelitian ini akan diuji optimasi terhadap model segmentasi berupa U-Net dan menggunakan model DenseNet-BC yang lebih efisien dalam jumlah parameter serta memiliki performa serupa dengan DenseNet.

3. Sistem yang Dibangun

3.1. Dataset

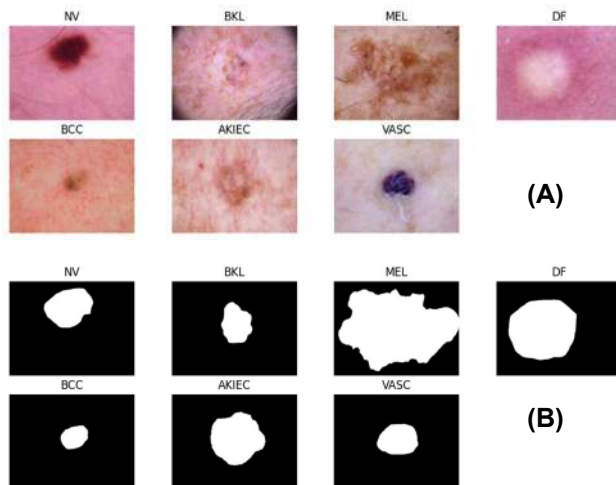
Dataset yang digunakan untuk penelitian ini adalah HAM10000, sebuah dataset dengan 7 label kategori berisi 10015 data berupa file foto dermoskopi dan *mask* foto tersebut. Masing-masing data memiliki satu dari 7 kategori label yaitu, *Actinic keratoses and intraepithelial carcinoma* (akiec), *basal cell carcinoma* (bcc), *benign keratosis-like lesions* (bkl), *dermatofibroma* (df), *melanoma* (mel), *melanocytic nevi* (nv), dan *vascular lesions* (vasc) dengan persebaran yang dapat dilihat pada Gambar 1. Dimensi citra dermoskopi dan *mask* pada dataset adalah 600x450 pixel.



GAMBAR 1

Persebaran data pada Dataset HAM10000 untuk setiap label.

Citra dermoskopi disimpan dengan format jpeg dan ruang warna RGB (*Red, Green, and Blue*) sedangkan *mask* disimpan dengan format png dan ruang warna *grayscale*, sampel data dari dataset dapat dilihat pada Gambar 2.

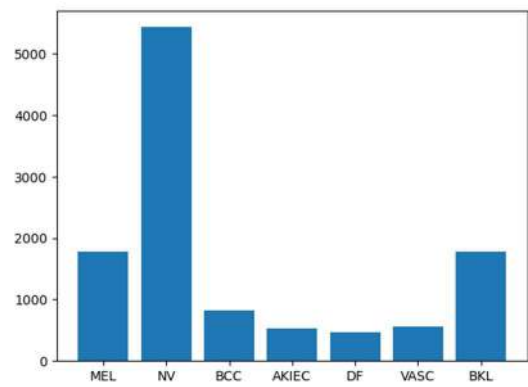


GAMBAR 2

Sampel masing-masing label dari dataset. (A) Sampel foto dermoskopi. (B) Sampel *mask*.

3.2. Praproses

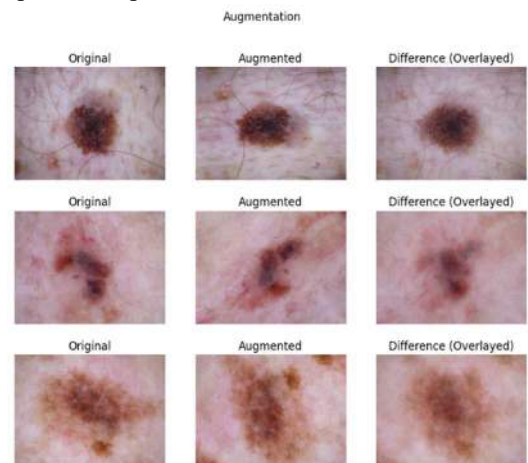
Praproses yang akan dilakukan berupa *resizing*, pembagian data, *oversampling*, augmentasi, dan normalisasi. Pertama *Input* akan di berlakukan *resizing* menjadi ukuran 300 x 400 pixel untuk mengurangi jumlah parameter model. Kemudian dataset dibagi menjadi 3, data latih, data validasi, dan data uji yang masing-masing memiliki 8014, 903, dan 1000 data. Berikutnya dilakukan *oversampling* pada data latih untuk mengurangi bias yang ada pada data, hasil *oversampling* dapat dilihat pada gambar 3.



GAMBAR 3

Persebaran data pada data latih setelah oversampling.

Kemudian augmentasi dilakukan pada data latih yang telah di *oversample* untuk mengurangi potensi overfitting pada model [10]. Augmentasi yang digunakan adalah *random horizontal and vertical flip*, dan *affine*, dalam bentuk rotasi acak dan *scaling* acak, dengan parameter bagi rotasi dan *scaling* adalah 0 hingga 360 dan 0,2, contoh augmentasi dapat dilihat pada Gambar 4.



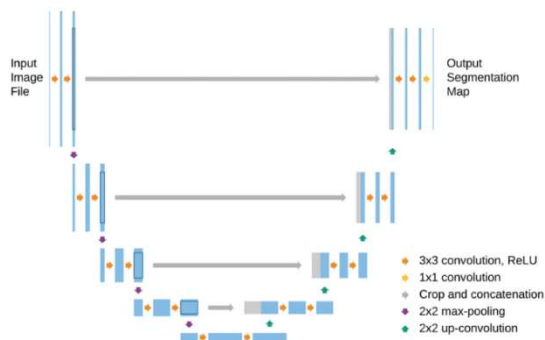
GAMBAR 4

Contoh hasil augmentas. Kolom pada bagian kiri menunjukkan citra tanpa augmentasi, kolom tengah menunjukkan citra setelah augmentasi, dan kolom pada bagian kanan untuk membantu visualisasikan perbedaan antara citra dengan augmentasi dan tanpa augmentasi.

Terakhir akan dilakukan normalisasi pada seluruh *input* baik dari data latih, validasi, dan uji. Normalisasi dilakukan untuk mengurangi efek dari perbedaan perangkat yang digunakan pada saat pengambilan data.

3.3. U-Net

U-Net adalah sebuah arsitektur berbasis CNN yang umumnya digunakan untuk tugas segmentasi. Arsitektur U-Net tersusun atas 2 bagian, bagian kontraksi atau *encoder* dan bagian ekspansi atau *decoder*, umumnya sebuah model U-Net memiliki jumlah layer kontraksi dan ekspansi yang serupa, menghasilkan arsitektur yang berbentuk simetris seperti huruf “U” [11], dapat dilihat pada Gambar 5.



GAMBAR 5

Diagram model U-Net, pada bagian kiri terdapat bagian encoder dan bagian kanan menunjukkan bagian decoder.

Model U-Net merupakan model yang efektif dalam melakukan segmentasi [12]. Model tersebut memiliki beragam variasi dan potensi modifikasi menjadikannya salah satu model paling fleksibel, yang memiliki sebuah varian untuk hampir segala kasus segmentasi. Pada penelitian ini, akan dipakai model U-Net dasar, dan optimasi dilakukan dengan merubah kedalaman atau jumlah pasangan *decoder* dan *encoder*.

3.4. DenseNet

Densely Connected Convolutional Network (DenseNet) merupakan sebuah arsitektur CNN yang memanfaatkan *dense connectivity* untuk mengambil *output* semua layer sebelumnya sebagai *input* dan mengeluarkan *output* yang akan diberikan ke seluruh layer setelahnya. *Dense connectivity* pada DenseNet membuat arsitektur ini membantu mengatasi gradient vanishing dan gradient exploding serta mengurangi potensi overfitting [13]. Penelitian ini memanfaatkan DenseBlock beserta Transition layer yang ada pada DenseNet dengan mengikuti konfigurasi DenseNet-BC (Bottleneck dan Compression). Konfigurasi DenseNet-BC dipilih untuk membantu meningkatkan efisiensi komputasi tanpa memberikan dampak yang terlalu besar pada performa model [14].

3.5. Kombinasi Model

Model yang dibangun pada penelitian ini berupa model kombinasi antara U-Net dan DenseNet masing-masing sebagai model segmentasi dan klasifikasi. Metode yang digunakan untuk menghubungkan kedua model sangat berpengaruh terhadap performa model classifier [15], maka akan diuji tiga buah metode, yaitu penghilangan latar belakang (*remove background*), *grayscale*, dan *4-channel*.

Metode penghilangan latar belakang merupakan metode sederhana untuk menggabungkan hasil segmentasi dengan input dengan merubah nilai pixel yang berada diluar *mask* menjadi nol, sehingga model dapat memfokuskan ekstraksi fitur pada

daerah penting, tanpa gangguan dari tekstur diluar *mask* seperti pori-pori atau pola pada kulit. *Grayscale* bekerja dengan menghitamkan nilai pixel diluar *mask*, kemudian merubah ruang warna citra menjadi abu-abu, perubahan ruang warna dipakai untuk menghilangkan variasi warna pada data, mengurangi sensitivitas model terhadap perubahan warna. Kemudian metode *4-channel* adalah metode sederhana yang menyematkan *mask* pada *channel* baru, metode ini memiliki keunggulan yaitu dapat menjaga fitur diluar *mask*, mengurangi dampak dari *mask* yang kurang akurat. Efek penggunaan metode masking terhadap input dapat dilihat pada gambar 6.



GAMBAR 6

Visualisasi metode *masking* yang digunakan. Untuk 4-channels, tidak ada perubahan yang dilakukan pada citra ataupun *mask*, hanya dilakukan penyematkan *mask* pada *channel* baru diatas *channel*-RGB.

3.6. Pelatihan Model

Pelatihan model akan dilakukan secara bertahap, dimulai dari model U-Net yang akan dilatih untuk segmentasi diikuti oleh model DenseNet yang telah dikombinasikan dengan U-Net. Pada saat pelatihan akan dilakukan pencarian konfigurasi optimal bagi U-Net, optimasi DenseNet, serta optimasi metode *masking* model.

Data latih yang digunakan untuk masing-masing model diacak menggunakan *seed* yang sama, dengan perbedaan untuk data latih segmentasi tidak dilakukan *oversampling* sedangkan pada data latih segmentasi akan dilakukan *oversampling*. Pada saat pelatihan model segmentasi, label yang digunakan berupa *mask*, dan untuk pelatihan klasifikasi akan digunakan label yang telah diberlakukan *one-hot-encoding*. Pelatihan masing-masing model akan dijalankan selama 20 epoch dengan fungsi *callback* yang menyimpan bobot terbaik model. Model segmentasi dengan bobot terbaik akan digunakan untuk melakukan pengujian pada model klasifikasi.

3.7. Metrik Evaluasi

Masing-masing model akan dievaluasi setelah pelatihan untuk mencari konfigurasi paling optimal. Evaluasi akan digunakan metrik berbasis *confusion matrix* yang mengidentifikasi hasil prediksi menjadi 4 kategori, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN), berdasarkan kebenaran prediksi terhadap *ground truth* dari label yang sedang dievaluasi, sehingga dapat diperoleh tabel kebenaran untuk masing-masing kategori seperti berikut.

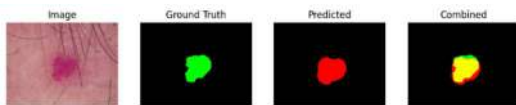
TABEL 1

Tabel confusion matrix untuk segmentasi dan klasifikasi. Nilai pixel pada segmentasi dapat bernilai 1 atau 0 untuk menunjukkan *lesion* atau *background* dengan label *lesion* merupakan label positif dan *background* berupa label negatif. Pada klasifikasi label dianggap apabila dilakukan

evaluasi terhadap kelas berlabel X sehingga label X menjadi label positif dan label selain X berupa label negatif.

Kategori Confusion Matrix	Segmentasi (value pixel)		Klasifikasi (untuk label X)	
	Prediksi	Ground	Prediksi	Ground
True Positive	1 (Lesion)	1 (Lesion)	X	X
True Negative	0 (Background)	0 (Background)	Bukan X	Bukan X
False Positive	1 (Lesion)	0 (Background)	X	Bukan X
False Negative	0 (Background)	1 (Lesion)	Bukan X	X

Confusion matrix umumnya digunakan mengevaluasi hasil klasifikasi sehingga untuk memperoleh nilai *confusion matrix* pada hasil segmentasi, akan diambil dari hasil prediksi masing-masing pixel terhadap pixel pada *ground truth*, untuk lebih jelas dapat dilihat pada gambar dibawah.



GAMBAR 7

Visualisasi identifikasi untuk masing-masing pixel terhadap *confusion matrix*. Gambar berlabel *Ground truth* berupa mask asli, dengan gambar berlabel *predicted* berupa mask hasil prediksi model. Gambar berlabel *Combined* merupakan gambar *Ground Truth* dengan *Predicted* yang ditindihkan pada satu gambar, pixel berwarna kuning menunjukkan True Positive, pixel berwarna hitam menunjukkan True Negative, pixel berwarna hijau menunjukkan False Negative, dan pixel berwarna merah menunjukkan False Positive.

Metrik evaluasi yang akan digunakan untuk segmentasi adalah *Mean IoU*, sensitivitas, *Specificity*, dan *F1-score*, sedangkan untuk klasifikasi akan digunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Rumus metrik- metrik yang dipakai adalah sebagai berikut:

$$\text{Akurasi} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Neg}} \quad (1)$$

$$\text{Sensitivitas/Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{Presisi} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (4)$$

$$F1\text{-Score/Dice} = 2 \times \frac{\text{True Positive}}{2(\text{True Positive}) + \text{False Positive} + \text{False Negative}} \quad (5)$$

$$\text{Mean IoU} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}} \quad (6)$$

Akurasi merupakan metrik yang paling umum digunakan, metrik tersebut menilai prediksi benar baik itu *true positive* atau *true negative* dibandingkan dengan keseluruhan prediksi, sedangkan metrik presisi merupakan metrik yang menilai hanya prediksi *true positive* dibandingkan dengan keseluruhan prediksi positif lainnya. Metrik sensitivitas dan *specificity* merupakan metrik yang umum digunakan di dunia medis untuk melakukan evaluasi. Metrik sensitivitas atau umumnya dikenal sebagai *recall* menunjukkan kemampuan model dalam memberikan suatu label dengan akurat yaitu *true positive* dibandingkan dengan *false positive*, sedangkan *specificity* menunjukkan kemampuan model saat mendeteksi sesuatu yang tidak termasuk label atau *true negative*. Kemudian metrik *mean IoU* dan *F1-score* atau *Dice Similarity Coefficient* (DSC), berupa metrik yang diperoleh dari sensitivitas dan presisi, yang keduanya menilai *overlap* dari hasil prediksi model dengan data asli, dengan DSC memberikan penilaian untuk performa rata-rata model tetapi *mean IoU* menilai performa terburuk model [16].

4. Evaluasi

Pengujian dilakukan dengan menggunakan layanan Kaggle dengan menggunakan seed yang sama untuk proses pengacakan pada proses pelatihan. Hyperparameter yang digunakan untuk semua model serupa, dengan menggunakan *optimizer* Adam dan *learning rate* sebesar 0.0001 serta *weight decay* 0.001. Untuk detail lebih lanjut terkait pengaturan *hyperparameter* dapat dilihat pada Tabel 2.

TABEL 2

Konfigurasi *hyperparameter* umum pada saat pelatihan.

Konfigurasi Hyperparameter				
Batch Size				128
Epoch				20
Optimizer				Adam
Learning Rate				0.0001
Weight Decay				0.001
U-Net				
Kedalaman	4			
Layer	Encoder		Decoder	
	Input	Output	Input	Output
Layer 1	3	64	64	1
Layer 2	64	128	128	64
Layer 3	128	256	256	128
Layer 4	256	512	512	256
Bottleneck	512			512
Kedalaman	5			
Layer	Encoder		Decoder	
	Input	Output	Input	Output

Layer 1	3	32	32	1
Layer 2	32	64	64	32
Layer 3	64	128	128	64
Layer 4	128	256	256	128
Layer 5	256	512	512	256
Bottleneck	512			512
Kedalaman	6			
Layer	Encoder		Decoder	
	Input	Output	Input	Output
Layer 1	3	16	16	1
Layer 2	16	32	32	16
Layer 3	32	64	64	32
Layer 4	64	128	128	64
Layer 5	128	256	256	128
Layer 6	256	512	512	256
Bottleneck	512			512

DenseNet – Kombinasi		
Faktor Bottleneck	2	
Faktor Kompresi	0.5	
Growth Rate	6	
Dropout Rate	0.5	
Layer	Input	Output
Konvolusi 1	X	256
Konvolusi 2	256	128
DenseBlock 1	128	512
Transition Layer 1	512	256
DenseBlock 2	256	512
Transition Layer 2	512	256
DenseBlock 3	256	512
Transition Layer 3	512	256
DenseBlock 4	256	1024
Transition Layer 4	1024	512
DenseBlock 5	512	1024
Transition Layer 5	1024	512
DenseBlock 6	512	1024
Output	1024	7
Metode Masking	Input Model (X)	
Tanpa Masking	3	
4-Channels	4	
Remove Background	3	
Grayscale	1	

4.1. Hasil Pengujian

Pengujian model U-Net dilakukan dengan variasi kedalaman 5 dan 6, selain itu model juga dilatih menggunakan 2 tipe dataset, dataset dengan normalisasi dan dataset tanpa normalisasi. Terdapat pula model yang dilatih dengan kedalaman 4 sebagai *baseline* yang diuji pada dataset tanpa normalisasi. Pelatihan model dilakukan dimulai dari dataset tanpa normalisasi dan kedalaman paling rendah, model kemudian diuji menggunakan bobot terbaik yang diraih saat pelatihan. Hasil pengujian adalah sebagai berikut.

TABEL 3

Hasil pengujian pada model Segmentasi U-Net. Konfigurasi nomor 1-3 dilatih dengan dataset yang tidak dinormalisasi, dengan konfigurasi nomor 1 berfungsi sebagai *baseline* menggunakan konfigurasi dasar U-Net.

Konfigurasi nomor 4-5 dilatih dengan dataset yang dinormalisasi.

No mo r	Aug men tasi	Nor mal isas i	Ke dal aman	m Io U	F 1 - S c o re	Sen siti vita s	Spe cifi city
1 (Ba seli ne)	Iya	Tid ak	4	5 2 %	5 8 %	54 %	99 %
2	Iya	Tid ak	5	6 7 %	7 4 %	72 %	97 %
3	Iya	Tid ak	6	8 7 %	9 2 %	93 %	96 %
4	Iya	Iya	5	8 7 %	9 2 %	94 %	97 %
5	Iya	Iya	6	8 7 %	9 2 %	92 %	97 %

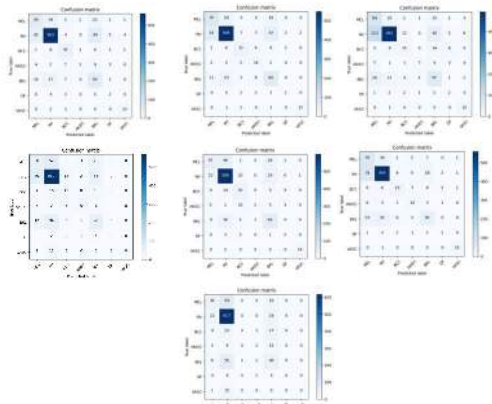
Berikutnya untuk pelatihan model DenseNet dan kombinasi akan diuji dengan menggunakan tiga metode *masking* berbeda, yaitu *4-channel*, penghilangan latar belakang, dan *grayscale*. Selain metode *masking*, model juga akan diuji dengan dataset dengan normalisasi dan tanpa normalisasi. Selanjutnya akan dilatih model DenseNet tanpa segmentasi yang berlaku sebagai *baseline*. Setiap konfigurasi model akan dilatih dengan jumlah *epoch* yang sama dan jumlah layer serta pengaturan optimizer yang sama. Hasil dari pengujian adalah sebagai berikut.

TABEL 4

Hasil evaluasi metode *masking*. M dan W berturut-turut melambangkan *Macro average* dan *Weighted average*. Nomor 1 merupakan *baseline* yang terdiri atas klasifier DenseNet saja. Model nomor 2-4 menggunakan hasil segmentasi U-Net nomor 3 (lihat Tabel 3), sedangkan model nomor 5-6 menggunakan hasil segmentasi U-Net nomor 4 (lihat Tabel 3).

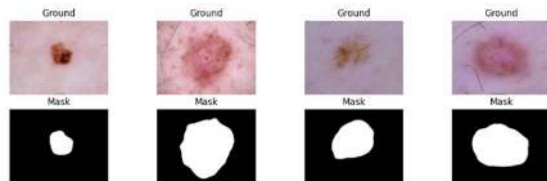
No mo r	Nor mali sasi	Me tod e Ma ski ng	Ak ur asi	Presi si		Recal l		F1- Score	
				M	W	M	W	M	W
1 (Ba seli ne)	Ya	-	73 %	5 3 %	7 5 %	5 3 %	7 3 %	5 0 %	7 4 %
2	Tida k	4- Ch ann els	74 %	5 5 %	7 6 %	5 7 %	7 4 %	5 5 %	7 5 %

3	Tidak	Remove bg.	67 %	47 %	74 %	52 %	67 %	46 %	69 %
4	Tidak	Grayscale	67 %	31 %	65 %	30 %	67 %	30 %	66 %
5	Iya	4-Channels	76 %	61 %	76 %	52 %	76 %	52 %	75 %
6	Iya	Remove bg.	73 %	53 %	75 %	43 %	73 %	53 %	73 %
7	Iya	Grayscale	70 %	38 %	67 %	40 %	72 %	55 %	65 %



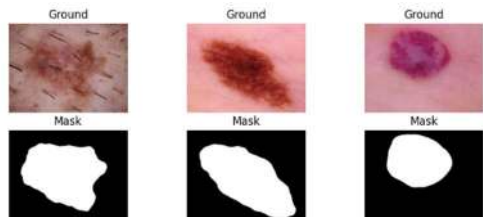
GAMBAR 8

Confusion matrix model. (1) Model baseline atau hanya DenseNet dengan normalisasi. (2) Model kombinasi metode 4-Channel tanpa normalisasi. (3) Model kombinasi metode remove background tanpa normalisasi. (4) Model kombinasi metode grayscale tanpa normalisasi. (5) Model kombinasi metode 4-Channel dengan normalisasi. (6) Model kombinasi metode remove background dengan normalisasi. (7) Model kombinasi metode grayscale dengan normalisasi.



Ground = AKIEC
Pred = NV

Ground = BCC
Pred = DF



Ground = MEL
Pred = MEL

Ground = NV
Pred = NV

GAMBAR 9

Sampel prediksi dari model terbaik (Nomor 5 pada tabel 4) untuk masing-masing label dalam dataset. Citra berlabel ground menunjukkan input model, citra berlabel mask menunjukkan hasil segmentasi dari U-Net terhadap ground, dan teks dibawah menunjukkan label dari input, dengan "Ground" menunjukkan label sebenarnya dan "Pred" label prediksi.

Untuk mengevaluasi pengaruh dari augmentasi dan oversampling pada dataset, model terbaik untuk dataset dengan normalisasi dan dataset tanpa normalisasi dibandingkan dengan model yang dilatih pada dataset tanpa augmentasi dan oversampling. Model yang dilatih berupa model U-Net dengan kedalaman 5 dan 6 dengan konfigurasi yang sama dengan tabel 2. Model U-Net terbaik akan dikombinasikan dengan densenet dengan konfigurasi seperti tabel 2. Model kombinasi yang akan dibandingkan adalah 2 model, menggunakan metode *masking 4-channels* dan *remove background*. Hasil evaluasi dapat dilihat pada tabel dibawah.

TABEL 5

Evaluasi pengaruh dari penggunaan metode augmentasi terhadap performa klasifikasi model. M dan W berturut-turut melambangkan *Macro average* dan *Weighted average*. Nomor 1 merupakan *baseline* yang terdiri atas klasifier DenseNet saja. Model nomor 1 menggunakan hasil segmentasi U-Net nomor 3 (lihat Tabel 3), sedangkan model nomor 2 menggunakan hasil segmentasi U-Net nomor 4 (lihat Tabel 3).

No.	Normalisasi	Augmentasi	Metode Masking	Akurasi	Presisi		Recall		F1-Score	
					M	W	M	W	M	W
1	Tidak	Ya	4-Channels	74 %	55 %	76 %	57 %	74 %	55 %	75 %
2	Ya	Ya	4-Channels	76 %	61 %	76 %	52 %	76 %	52 %	75 %
3	Tidak	Tidak	Remove bg	74 %	50 %	73 %	47 %	74 %	48 %	73 %
4	Tidak	Tidak	4-Channels	76 %	55 %	73 %	45 %	76 %	47 %	73 %

4.2. Analisis Hasil Pengujian

Dari pengujian model segmentasi, konfigurasi model nomor 1 meskipun berhasil mencapai nilai spesifisitas paling tinggi yaitu 99%, konfigurasi ini memiliki nilai mIoU, *F1-Score*, dan sensitivitas paling rendah, sebesar 52%, 58%, dan 54% berturut-turut. Konfigurasi nomor 3,4, dan 5 mencapai nilai mIoU dan F1-Score serupa, tetapi konfigurasi nomor 4 berhasil mencapai nilai rata-rata tertinggi dari semua konfigurasi yang diuji yaitu *mean IoU*

sebesar 87%, *F1-Score* rata-rata sebesar 92%, sensitivitas 94%, dan *specificity* 97%.

Berikutnya pada hasil evaluasi model kombinasi pada Tabel 4, terdapat perbedaan yang sangat drastis dari performa model yang digabungkan dengan menggunakan metode *4-Channel* dan *remove background*/penghilangan latar belakang, dengan metode *grayscale*, menandakan adanya hubungan erat antara label dengan warna pada lesi. Hal ini dapat menjelaskan mengapa normalisasi dapat meningkatkan performa model dengan baik. Ketergantungan terhadap ruang warna pula dapat menjelaskan mengapa metode *grayscale* tidak dapat melakukan klasifikasi untuk 2 label, yaitu AKIEC dan DF, hal ini dapat dilihat pada gambar 8, matrix nomor 4 dan 7.

Selain itu metode penghilangan latar belakang memiliki performa yang hampir sama dengan model 1. Pada dasarnya metode penghilangan latar belakang bertujuan untuk menghilangkan gangguan tekstur diluar *mask*, namun apabila *mask* yang digunakan memiliki sensitivitas atau *specificity* yang rendah, metode ini dapat menghilangkan fitur penting dari citra. Untuk model kombinasi 3, digunakan model segmentasi yang memiliki sensitivitas dan *specificity* yang sedikit lebih rendah dibandingkan dengan model segmentasi pada model kombinasi 6. Untuk menghindari kemunculan masalah serupa, dapat ditambahkan sedikit dilasi pada saat menghilangkan latar belakang untuk mengantisipasi *mask* yang kurang baik.

Setelah melakukan pengujian, hanya metode kombinasi *4-Channels* yang dengan konsisten dapat memperoleh hasil diatas model baseline, meskipun hanya 3%, dengan metode lain memperoleh hasil lebih buruk ataupun hanya setara dengan model baseline.

Penggunaan normalisasi sangat efektif meningkatkan performa model, dengan hampir seluruh model yang dijalankan menggunakan dataset yang telah di normalisasi memperoleh performa yang lebih baik. Model nomor 2 dan 4 hanyalah dua model yang dapat memperoleh performa lebih baik atau setara pada paling tidak salah satu metrik evaluasi yang digunakan. Dari seluruh metode *masking* yang digunakan, metode *grayscale* memperoleh hasil terburuk secara keseluruhan. Metode augmentasi yang digunakan apabila digunakan bersamaan dengan normalisasi memiliki hasil yang lebih baik dibandingkan dengan model yang dilatih tanpa normalisasi dan augmentasi, tetapi apabila augmentasi saja yang digunakan performa model cenderung menurun dibandingkan dengan yang dilatih tanpa augmentasi. Apabila dilihat dari gambar 9, prediksi yang salah terjadi karena ketidakmampuan model klasifikasi dalam membedakan detail yang kecil antara masing-masing label. Bagian segmentasi dari model dapat membuat *mask* yang mencakup hampir seluruh bagian dari lesi, namun kemiripan dari bentuk lesi antar label (label AKIEC pada gambar 9 dan label

NV pada gambar 2, label BCC dan label DF pada gambar 9) menunjukkan kalau kelemahan model terletak pada bagian klasifikasi. Tetapi untuk label seperti VASC yang memiliki fitur yang cukup jelas seperti warna merah darah, model tidak memiliki kesulitan dalam mendeteksi label tersebut.

Sistem memperoleh hasil paling optimal dengan menggunakan dataset yang telah di normalisasi, dengan metode *masking* model yang tidak berpotensi destruktif terhadap fitur penting. Metode *grayscale* dan penghilangan latar belakang gagal memperoleh hasil efektif karena kedua metode memiliki sebuah sifat destruktif baik dalam bentuk menghilangkan fitur ataupun mengurangi ruang warna. Model terbaik yang diraih pada penelitian ini mencapai nilai akurasi 76%, dengan presisi rata-rata berbobot sebesar 76%, *recall* rata-rata berbobot 76%, dan *F1-Score* rata-rata berbobot sebesar 75%.

5. Kesimpulan dan Saran

Penelitian ini berhasil menyusun pipeline yang baik untuk membangun model kombinasi klasifikasi dengan segmentasi. Model pertama melakukan segmentasi terhadap *input* yang menghasilkan sebuah *mask*, yang kemudian akan digunakan untuk melakukan *masking* kepada data *input* dan data yang telah di *mask* akan digunakan sebagai *input* bagi model klasifikasi. Model segmentasi pertama dioptimasi dengan mencari konfigurasi yang dapat menghasilkan evaluasi terbaik, yang kemudian digunakan untuk menguji metode *masking* yang dapat diimplementasikan. Meskipun model akhir hanya mencapai akurasi 76% dengan presisi rata-rata berbobot 76%, *recall* rata-rata berbobot 76%, dan *F1-Score* rata-rata berbobot 75%, model tersebut masih memperoleh hasil yang lebih baik dibandingkan penggunaan model secara individu pada lingkungan yang sama.

Untuk penelitian kedepan, dapat diuji dengan menggunakan model segmentasi atau klasifikasi lainnya, atau melakukan eksplorasi lebih dalam pada parameter model yang dipakai, karena terdapat beberapa parameter yang tidak bisa diuji dalam penelitian ini dikarenakan keterbatasan sistem yang digunakan. Selain itu dapat pula diuji metode *masking* lainnya yang tidak dapat diimplementasikan pada penelitian ini dikarenakan konflik dengan implementasi model.

Daftar Pustaka

- [1] E. R. Parker, "The influence of climate change on skin cancer incidence – A review of the evidence," *International Journal of Women's Dermatology*, pp. 17-27, 2021.
- [2] R. Gordon, "Skin Cancer: An Overview of Epidemiology and Risk Factors," *Seminars in Oncology Nursing*, vol. 29, no. 3, pp. 160-169, 2013.
- [3] G. Liezel L, L. John T and A. Faisal Rehman, "Non-melanoma skin cancer," *Clinical Medicine*, vol. 16, no. 1, pp. 62-65, 2016.

- [4] H. Md. Kamrul, A. Md. Asif, Y. Choon Hwai and Y. Guang, "A survey, review, and future trends of skin lesion segmentation and classification," *Computers in Biology and Medicine*, vol. 155, p. 106624, 2023.
- [5] A. A. Adegun and S. Viriri, "FCN-Based DenseNet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images," *IEEE Access*, vol. 8, pp. 150377-150396, 2020.
- [6] M. Dildar, S. Akram, M. Irfan, H. U. Khan, M. Ramzan, A. R. Mahmood, S. A. Alsaiani, A. H. M. Saeed, M. O. Alraddadi and M. H. Mahnashi, "Skin Cancer Detection: A Review Using Deep Learning Techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, 2021.
- [7] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," *Artificial Intelligence in Medicine*, vol. 102, p. 101756, 2020.
- [8] 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*, no. 1, 2018.
- [9] V. Anand, S. Gupta, D. Koundal and K. Singh, "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images," *Expert Systems with Applications*, vol. 213, p. 119230, 2023.
- [10] F. Perez, C. Vasconcelos, S. Avila and E. Valle, "Data Augmentation for Skin Lesion Analysis," Cham, 2018.
- [11] N. Siddique, S. Paheding, C. P. Elkin and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access*, vol. 9, pp. 82031-82057, 2021.
- [12] W. Jiangtao, N. I. R. Ruhaiyem and F. Panpan, "A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation," *IET Image Processing*, vol. 19, p. e70019, 2025.
- [13] T. Li, W. Jiao, L.-N. Wang and G. Zhong, "Automatic DenseNet Sparsification," *IEEE Access*, vol. 8, pp. 62561-62571, 2020.
- [14] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [15] A. Mahbod, P. Tschandl, G. Langs, R. Ecker and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105725, 2020.
- [16] D. Müller, I. Soto-Rey and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.
- [17] H. Basak, R. Kundu and R. Sarkar, "MFSNet: A multi focus segmentation network for skin lesion segmentation," *Pattern Recognition*, vol. 128, p. 108673, 2022.