

Perbandingan Model U-Net dan ELU-Net untuk Segmentasi Semantik Citra Medis Kanker Pankreas

Algi Fari Ramdhani¹, Yudi Widhiyana¹, Setiadi Rachmat¹

¹ Program Studi Teknik Informatika, Jurusan Teknik Komputer dan Informatika, Politeknik Negeri Bandung, Kabupaten Bandung Barat, Jawa Barat 40559, Indonesia

[Diserahkan: 23 September 2024, Direvisi: 13 Desember 2024, Diterima: 3 Februari 2025]
Penulis Korespondensi: Algi Fari Ramdhani (email: algi.fari.tif420@polban.ac.id)

INTISARI — Analisis citra medis untuk melakukan segmentasi semantik menggunakan teknologi *deep learning* masih dikembangkan saat ini. Salah satu pengembangannya adalah arsitektur U-Net, yang dapat menghasilkan akurasi yang baik. Pengembangan dilanjutkan menghasilkan ELU-Net dengan fokus membuat model makin efisien. ELU-Net menghasilkan akurasi yang cukup baik, tetapi perlu adanya kajian lebih lanjut mengenai perbandingan kedua model. Kedua model akan dibandingkan berdasarkan akurasi, penggunaan penyimpanan, dan waktu proses dalam melakukan segmentasi semantik citra kanker pankreas. Citra kanker pankreas yang digunakan pada penelitian ini berasal dari sebuah PAIP 2023 Challenge yang berisi sebuah citra dengan pewarnaan *haematoxylin* dan *eosin* (H&E). Eksperimen dilakukan dengan mengubah jumlah filter dan kedalaman model untuk kedua arsitektur. Evaluasi dilakukan terhadap *dataset* citra kanker pankreas yang berjumlah 57 citra. Dari serangkaian percobaan yang dilakukan, diperoleh hasil bahwa U-Net memiliki akurasi terbaik sebesar 92,8%, sedikit lebih unggul dibandingkan ELU-Net yang mencapai 89,7%. Namun, ELU-Net lebih efisien dari segi penggunaan penyimpanan (8,1 MB untuk ELU-Net dan 93,31 MB untuk U-Net) dan waktu proses (4,0 s untuk ELU-Net dan 5,3 s untuk U-Net). Akurasi yang diperoleh ELU-Net memang lebih kecil daripada U-Net, tetapi dari penggunaan penyimpanan dan waktu proses, ELU-Net jauh unggul dengan selisih 85,21 MB dan 1,3 detik. Dari hasil tersebut, model ELU-Net tidak lebih baik daripada U-Net, terutama dari segi akurasi. Namun, dengan perbandingan ELU-Net terhadap U-Net pada ukuran penyimpanan sebesar 1:11,51 dan waktu proses 1:1,325, selisih akurasi 3,1% merupakan *trade-off* yang cukup logis.

KATA KUNCI — U-Net, ELU-Net, *Lightweight Model*, Segmentasi Semantik.

I. PENDAHULUAN

Menurut World Health Organization (WHO), kanker adalah salah satu penyakit dengan tingkat kematian yang sangat tinggi di seluruh dunia [1]. Berbagai jenis kanker menimbulkan ancaman serius terhadap kesehatan global dan penanganannya memerlukan perhatian serta intervensi khusus. Salah satu jenis kanker yang sangat mematikan adalah kanker pankreas, yang dikenal menyebabkan pasien memiliki tingkat kelangsungan hidup yang sangat rendah, yakni hanya sekitar 10% [2]. Ini berarti bahwa hanya sedikit pasien yang dapat bertahan hidup lebih dari lima tahun setelah diagnosis awal. Salah satu faktor utama penyebab rendahnya angka kelangsungan hidup ini adalah kurangnya penanganan yang memadai, terutama dalam hal deteksi dini.

Deteksi dini pada kanker pankreas memiliki peranan yang sangat krusial dalam meningkatkan peluang bertahan hidup pasien [3]. Kanker pankreas sering kali tidak menunjukkan gejala pada tahap awal, sehingga sering kali didiagnosis ketika sudah mencapai tahap lanjut, yang menyebabkan pengobatan menjadi jauh lebih sulit dan kurang efektif. Dengan deteksi dini, kanker pankreas dapat ditemukan pada tahap yang lebih awal, memungkinkan pengobatan yang lebih tepat waktu dan efektif. Upaya untuk meningkatkan kesadaran akan pentingnya deteksi dini dan pengembangan teknologi serta metode diagnostik yang lebih baik sangat diperlukan untuk meningkatkan angka kelangsungan hidup pasien kanker pankreas.

Deteksi kanker merupakan salah satu langkah krusial dalam proses pengobatan kanker secara keseluruhan. Deteksi yang akurat dan tepat waktu memungkinkan dokter menentukan diagnosis yang sesuai, yang kemudian dapat menjadi dasar bagi strategi pengobatan yang efektif. Salah satu aspek paling penting dalam deteksi kanker adalah deteksi dini. Hal ini

berpengaruh kepada keberhasilan dalam mengidentifikasi kanker pada tahap awal dan secara signifikan meningkatkan peluang pasien untuk sembuh. Namun, meskipun penting, pelaksanaan deteksi dini dalam praktik sehari-hari sering kali dihadapkan pada berbagai hambatan yang kompleks. Salah satu hambatan utama dalam deteksi dini kanker adalah proses ini masih banyak dilakukan secara manual oleh ahli patologi [4], [5]. Deteksi manual ini menimbulkan beberapa tantangan, terutama terkait dengan isu subjektivitas dan keterbatasan kemampuan manusiawi dari para ahli patologi. Subjektivitas muncul karena analisis citra medis bergantung pada pengetahuan dan pengalaman individu dari setiap ahli patologi. Dua ahli patologi yang menganalisis citra yang sama bisa saja menghasilkan diagnosis yang berbeda, tergantung pada tingkat pengalaman dan interpretasi terhadap citra tersebut.

Selain subjektivitas, keterbatasan lain yang signifikan adalah keterbatasan fisik dan mental dari ahli patologi. Seperti halnya pekerjaan lain yang membutuhkan ketelitian dan konsentrasi tinggi, ahli patologi juga tidak dapat bekerja secara terus-menerus tanpa mengalami kelelahan [6]. Pekerjaan yang berjam-jam memeriksa citra medis yang kompleks dapat menyebabkan penurunan konsentrasi dan akurasi, yang pada akhirnya dapat memengaruhi kualitas hasil analisis. Faktor kelelahan ini bisa menjadi lebih kritis dalam kasus tertentu, seperti sangat banyaknya jumlah kasus yang harus ditangani oleh seorang patologi, sehingga meningkatkan risiko kesalahan diagnosis.

Untuk dapat mengatasi deteksi kanker yang masih manual diperlukan pendekatan lain, salah satunya adalah dengan menggunakan teknologi [7]. Pengembangan teknologi untuk membantu melakukan analisis deteksi dini adalah dengan menggunakan *deep learning* [8]. Jenis dari *deep learning* yang

khusus untuk analisis sebuah citra adalah *convolutional neural-network* (CNN). Penelitian pengembangan CNN sedang masif dilakukan pada saat ini, terutama dalam analisis citra medis [9]. Analisis citra dengan CNN juga menghasilkan akurasi yang cukup bagus.

Dalam melakukan analisis, CNN memerlukan sebuah citra medis. Salah satu analisis yang dapat dilakukan oleh CNN adalah segmentasi semantik. CNN akan dibuat berdasarkan arsitektur tertentu untuk dapat melakukan segmentasi semantik.

Salah satu jenis CNN yang dapat melakukan segmentasi semantik adalah U-Net [10]. U-Net khusus dibuat untuk melakukan segmentasi semantik pada citra medis. Pertama kali U-Net dibuat untuk melakukan segmentasi semantik pada citra dengan pewarnaan *differential interference contrast* (DIC) pada sebuah sel HeLa. Hasil akurasi U-Net dalam melakukan segmentasi semantik cukup baik dibandingkan model lain pada jenis data penelitian yang sama.

Akurasi yang tinggi telah dicapai oleh U-Net. Namun, untuk mencapainya, model ini membutuhkan sumber daya yang cukup besar [11]. Lebih tepatnya, model U-Net memerlukan sekitar 31 juta parameter, yang mencerminkan kompleksitas dan kebutuhan komputasi yang signifikan. Jumlah parameter yang besar ini memungkinkan U-Net untuk menangkap detail yang lebih akurat dalam proses segmentasi, sehingga ini bisa menjadi kelebihan dari model U-Net. Namun, di sisi lain, hal ini juga membuat model U-Net membutuhkan memori yang lebih banyak dan waktu eksekusi yang lebih lama. Hal ini mengakibatkan model U-Net kurang efisien dalam beberapa aplikasi, terutama ketika dihadapkan dengan penggunaan di dunia nyata.

Dalam upaya untuk mengembangkan model yang lebih ringan, tetapi tetap efektif, lahirlah ELU-Net. ELU-Net merupakan pengembangan dari U-Net yang tidak hanya mempertahankan kemampuan segmentasi yang baik, tetapi juga mengurangi jumlah parameter yang dibutuhkan secara drastis [12]. Salah satu inovasi pada ELU-Net adalah penambahan *deep skip connection*, yang memungkinkan model memproses informasi dari citra dengan lebih efisien. Pada penelitian tersebut, U-Net diintegrasikan dengan VGG16 pada bagian *encoder* yang menghasilkan model dengan jumlah parameter sebanyak 600 ribu. Hal ini merupakan pengurangan jumlah parameter yang signifikan dibandingkan dengan U-Net, yaitu sebanyak 31 juta.

VGG16 sendiri adalah arsitektur *deep learning* yang awalnya dirancang untuk tugas klasifikasi citra dan telah terbukti sangat efektif dalam menangani berbagai jenis data citra [13]. Dengan menggabungkan VGG16 ke dalam ELU-Net, model ini berhasil mencapai keseimbangan antara akurasi tinggi dan efisiensi parameter, memungkinkan penggunaan yang lebih luas dalam lingkungan dengan sumber daya terbatas.

Meskipun memiliki jumlah parameter yang lebih sedikit, model dengan arsitektur ELU-Net tidak menunjukkan penurunan kinerja dibandingkan U-Net dari segi akurasi. Sebaliknya, dalam beberapa penelitian, ELU-Net dengan VGG16 bahkan mampu mencapai akurasi sebesar 0,96 berdasarkan evaluasi menggunakan metrik akurasi koefisien Dice. Hasil evaluasi ELU-Net ini lebih tinggi dibandingkan U-Net yang hanya mencapai akurasi 0,92. Koefisien Dice adalah salah satu metrik evaluasi yang sering digunakan dalam penelitian segmentasi citra, selain indeks Jaccard, karena mampu mengukur kesesuaian antara hasil prediksi dengan *ground truth* secara efektif [14].

Hasil yang mengesankan dari ELU-Net ini diperoleh dari data penelitian yang melibatkan citra organ otak, yang menyebabkan ELU-Net mampu memberikan hasil segmentasi yang lebih akurat dibandingkan dengan U-Net. Meskipun demikian, untuk memvalidasi keunggulan ELU-Net secara lebih luas, diperlukan perbandingan lebih lanjut dengan menggunakan berbagai jenis citra medis lainnya. Ini penting untuk memastikan bahwa keunggulan ELU-Net dalam hal akurasi dan efisiensi benar-benar konsisten di berbagai kondisi dan jenis data. Maka, pada penelitian ini, model ELU-Net dan U-Net akan dibandingkan dengan menggunakan citra dari PAIP 2023 Challenge. *Challenge* ini menyediakan citra kanker pankreas yang telah diwarnai dengan *hematoxylin* dan *eosin* (H&E). U-Net telah terbukti efektif dalam melakukan segmentasi semantik pada citra H&E [15], sehingga perbandingan ini akan memberikan wawasan penting mengenai kinerja relatif antara U-Net dan ELU-Net dalam konteks spesifik ini.

Dalam penelitian ini, kinerja kedua model tidak hanya dievaluasi berdasarkan akurasi segmentasi, tetapi juga dianalisis dari segi jumlah parameter yang digunakan, ukuran *file* model, dan waktu eksekusi pada tahap *inference*. Evaluasi menyeluruh ini akan memberikan gambaran yang lebih komprehensif tentang kelebihan dan kekurangan masing-masing arsitektur serta potensi penerapannya dalam lingkungan klinis yang membutuhkan keseimbangan antara akurasi dan efisiensi.

II. METODOLOGI

Bagian ini membahas hal-hal terkait penelitian, yaitu data penelitian, pemrosesan citra, pelatihan model, dan pengujian model.

A. DATA PENELITIAN

Penelitian ini menggunakan data yang sudah disediakan oleh penyelenggara PAIP 2023 Challenge. Untuk mendapatkan data, perlu dilakukan pendaftaran terlebih dahulu untuk mengikuti *challenge* dan pengisian formulir untuk mengakses data. Pendaftaran *challenge* dilakukan pada laman web <https://2023paip.grand-challenge.org/>. Setelah mendaftar *challenge*, dilakukan pendaftaran akun ke halaman <http://wisepaip.org/challenge2023>, dengan menggunakan surel yang sama pada halaman *challenge*. Kemudian, data diunduh sesuai dengan kebutuhan. PAIP 2023 Challenge adalah seri *challenge* pertama dari PAIP yang menggunakan *dataset* berupa citra kanker atau tumor pankreas.

Spesifikasi data dari penyelenggara *challenge* ditampilkan pada Tabel I. Terdapat 53 data pelatihan beserta label dengan 50 citra pankreas dan 3 citra usus besar. Sejumlah 53 data tersebut memiliki ukuran atau dimensi yang sama, yaitu 1.024×1.024 . Terdapat citra dengan perbesaran 0,250 mpp (mikron per piksel) dan 0,500 mpp. Mayoritas citra adalah citra pankreas dengan perbesaran 0,250 mpp, sehingga pada penelitian ini data yang digunakan adalah citra pankreas dengan perbesaran 0,250 mpp.

Data validasi dan pengujian dari PAIP 2023 Challenge tidak digunakan dalam penelitian ini karena kedua *dataset* tersebut tidak memiliki label anotasi yang diperlukan untuk tahapan pelatihan model. Tanpa label, data tidak dapat digunakan untuk pelatihan atau validasi kinerja model secara efektif karena label diperlukan untuk mengukur kesesuaian prediksi model terhadap *ground truth*. Oleh karena itu, penelitian ini hanya memanfaatkan data pelatihan yang

TABEL I
SPESIFIKASI DATA PAIP 2023

Jenis Data	Ukuran Citra (piksel)	Jumlah Citra Pankreas	Jumlah Citra Usus Besar
Latih		50	3
Validasi	1.024 × 1.024	10	-
Pengujian		20	20
Total		80	3

disediakan oleh PAIP 2023 Challenge, yang terdiri atas citra pankreas dengan resolusi perbesaran 0,250 mpp. Data pelatihan ini sudah dilengkapi dengan label yang memungkinkan model belajar mengenali pola-pola spesifik yang relevan dengan tugas segmentasi citra pankreas. Spesifikasi lengkap mengenai data yang digunakan dalam penelitian ini, termasuk jumlah citra dan ukuran citra, disajikan pada Tabel II.

Salah satu contoh citra yang digunakan pada penelitian ini ditunjukkan pada Gambar 1 dan Gambar 2. Gambar 1 merupakan citra H&E dari pasien yang didiagnosis mengalami kanker pankreas. Terlihat bahwa terdapat titik-titik yang berwarna lebih gelap, yang merupakan sebuah sel. Namun, tidak semua sel yang berwarna gelap merupakan sel kanker. Hal ini dapat dilihat pada Gambar 2, yang merupakan label anotasi dari Gambar 1. Warna putih pada Gambar 2 merupakan bagian-bagian yang merupakan sel kanker, sedangkan warna hitam merupakan *background* atau sel yang bukan kanker.

Sesuai dengan Tabel II, data pelatihan dan validasi berjumlah 36. Data pelatihan dan validasi diacak pada tahapan pelatihan untuk menghindari *overfitting*. Sementara itu, untuk pengujian digunakan sepuluh citra. Selanjutnya, pada 46 citra ini dilakukan prapemrosesan citra.

B. PRAPEMROSESAN CITRA

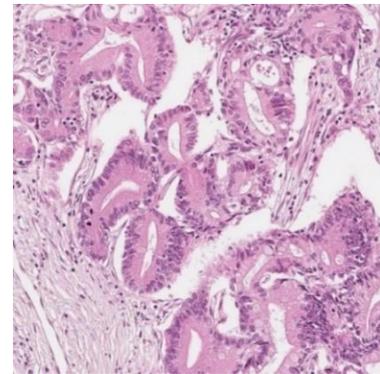
Dalam penelitian ini, tahap prapemrosesan citra dilakukan dengan tujuan utama untuk meningkatkan kualitas citra. Proses ini akan memengaruhi akurasi model yang dilatih. Kualitas citra yang baik dapat membantu model dalam mengenali pola-pola penting, sehingga meningkatkan kinerja model dalam segmentasi citra medis. Salah satu teknik prapemrosesan citra yang umum digunakan adalah normalisasi citra. Normalisasi ini bisa dilakukan dengan berbagai metode, salah satunya adalah teknik *histogram equalization* [16]. Teknik ini bertujuan untuk meningkatkan kontras citra dengan mendistribusikan intensitas piksel secara lebih merata, sehingga detail pada citra menjadi lebih menonjol. Proses *histogram equalization* dalam penelitian ini dilakukan menggunakan *library* OpenCV. OpenCV merupakan *library* pemrosesan citra yang sering digunakan dalam bidang *computer vision*.

Data yang digunakan untuk pelatihan model dalam penelitian ini tergolong sedikit, yaitu hanya terdiri atas 36 buah citra. Data ini digunakan untuk pelatihan dan validasi. Pembagian citra menjadi data pelatihan dan validasi dilakukan secara acak. Pembagian ini menerapkan perbandingan jumlah data pelatihan dan validasi sebesar 80:20. Dengan demikian, dari 36 citra, sekitar 29 citra digunakan untuk pelatihan dan 7 citra lainnya untuk validasi. Karena keterbatasan jumlah data, dilakukan *image augmentation* untuk memperbanyak jumlah data yang tersedia.

Image augmentation pada penelitian ini dilakukan dengan menerapkan beberapa transformasi sederhana, yaitu *flip* secara vertikal, horizontal, dan kombinasi keduanya [17]. Dengan menerapkan transformasi ini, setiap citra asli menghasilkan tiga citra baru setelah proses *augmentation*, sehingga jumlah total

TABEL II
SPESIFIKASI DATA PENELITIAN

Jenis Data	Ukuran Citra (piksel)	Jumlah Citra Pankreas
Latih		36
Validasi	1.024 × 1.024	10
Pengujian		46
Total		



Gambar 1. Citra H&E pada pankreas.



Gambar 2. Label dari Gambar 1.

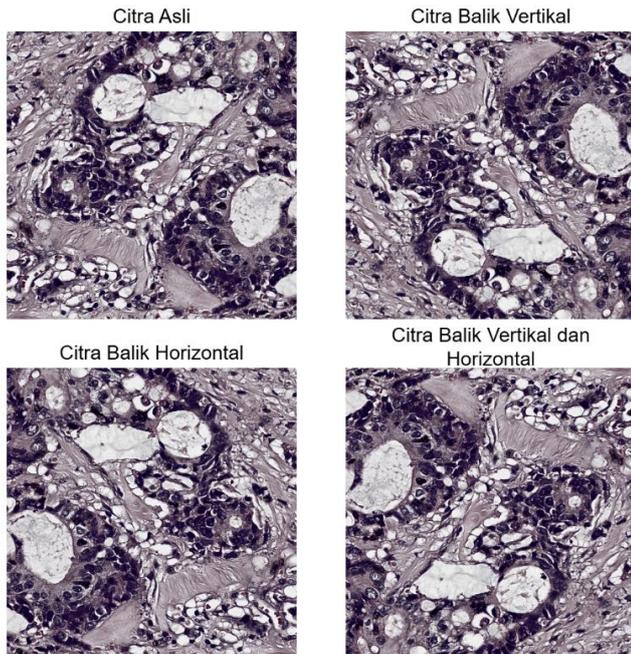
citra meningkat dari 36 menjadi 144 citra. Setiap citra yang telah mengalami proses *augmentation* memiliki dimensi 1.024 × 1.024 piksel, sama seperti citra asli. Contoh hasil dari proses *image augmentation* ditunjukkan pada Gambar 3.

Selanjutnya, sebuah citra dengan dimensi 1.024 × 1.024 dipotong menjadi 16 bagian yang lebih kecil, sehingga setiap bagian memiliki dimensi 256 × 256 piksel. Potongan-potongan citra ini kemudian digunakan sebagai data pelatihan atau validasi. Sebelum digunakan dalam pelatihan model, citra-citra tersebut diacak urutannya dan diubah menjadi sebuah larik NumPy. Proses pengubahan menjadi larik NumPy dilakukan agar citra-citra tersebut dapat menjadi masukan yang sesuai untuk model yang akan dilatih. Dengan demikian, setiap citra siap untuk menjadi data pelatihan model U-Net dan ELU-Net.

C. PELATIHAN MODEL

Dalam penelitian ini, pelatihan model mencakup beberapa tahapan penting, mulai dari pembuatan model dengan menggunakan *library* Keras, pelatihan model, hingga menyimpan model yang telah dilatih. Penyiapan model dilakukan dengan merancang berbagai variasi model, yang masing-masing diuji untuk menentukan konfigurasi terbaik. Variasi model-model yang dibuat mencakup kedalaman model dan jumlah filter awal pada proses konvolusi, sehingga setiap jenis model akan mendapatkan variasi yang sama.

Kedalaman model merupakan atribut yang krusial karena pada penelitian ELU-Net ditemukan bahwa kedalaman 5 pada



Gambar 3. Image augmentation dengan flip.

ELU-Net bukanlah patokan bahwa kedalaman tersebut merupakan yang terbaik [12]. Hal ini mendorong dilakukannya penelitian lebih lanjut untuk memastikan bahwa kedalaman 5 memang merupakan kedalaman yang optimal atau ada kedalaman lain yang lebih baik. Sementara itu, jumlah filter awal diambil sebagai atribut tambahan karena pada penelitian lain disebutkan bahwa model U-Net dengan jumlah filter awal yang lebih rendah dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan filter awal yang lebih besar [11].

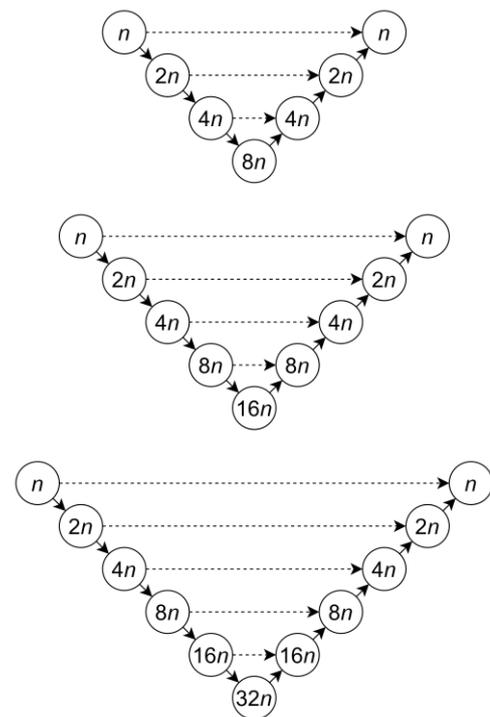
Pada arsitektur U-Net, kedalaman model yang diusulkan adalah 5 [10], dengan filter awal konvolusi sebesar 64. Sebaliknya, pada arsitektur ELU-Net, ditetapkan kedalaman model sebesar 5 dengan filter awal sebesar 8 [12]. Mengingat hasil-hasil penelitian ini, untuk menentukan model terbaik yang menggunakan data penelitian dari PAIP 2023 Challenge, diperlukan pembuatan beberapa versi model dengan variasi kedalaman dan jumlah filter awal. Deskripsi lebih lengkap mengenai versi model yang diuji disajikan pada Tabel III.

Kedalaman model yang dimaksud dalam penelitian ini merujuk pada banyaknya tahapan *down-sampling* yang dilakukan pada bagian *contractive path* dari model. Dalam konteks arsitektur U-Net, kedalaman ini menentukan jumlah lapisan konvolusi yang ada di setiap kedalaman, sebelum dilakukan *up-sampling* pada *expansive path*. Contoh dari beberapa kedalaman pada arsitektur U-Net ditunjukkan pada Gambar 4; secara berturut-turut dari atas ke bawah, arsitektur ditampilkan dengan kedalaman 4, 5, dan 6. Pada Gambar 4, nilai n menunjukkan jumlah filter awal pada setiap tahapan konvolusi. Maka, jumlah filter pada konvolusi terbesar ada pada kedalaman yang terdalam, yaitu sebanyak $8n$ untuk kedalaman 4, $16n$ untuk kedalaman 5, dan $32n$ untuk kedalaman 6.

Proses pelatihan model dilakukan menggunakan Kaggle Notebooks, yang menyediakan (*graphics processing unit*) GPU T4 untuk mempercepat pelatihan model. Sebelum pelatihan dimulai, model perlu di-*compile* terlebih dahulu. Pada tahapan ini, beberapa konfigurasi model ditambahkan. Konfigurasi model mencakup penambahan *optimizer* Adam dengan nilai laju pembelajaran sebesar 1×10^{-4} [18], penggunaan fungsi *loss binary cross entropy* [12], dan metrik akurasi untuk

TABEL III
 VARIASI MODEL U-NET DAN ELU-NET

Jenis Arsitektur	Jumlah Filter	Jumlah Kedalaman
U-Net	32	4
		5
		6
	64	4
		5
		6
ELU-Net	128	4
		5
		6
	4	4
		5
		6
8	4	4
		5
		6
	16	4
		5
		6



Gambar 4. Contoh jumlah kedalaman arsitektur U-Net.

mengevaluasi kinerja model. Setelah model di-*compile*, langkah selanjutnya adalah melakukan pelatihan model dengan konfigurasi yang telah disiapkan.

Pada tahap pelatihan, beberapa atribut penting digunakan, termasuk ukuran *batch* sebesar 8, *epoch* sebanyak 40, *validation split* sebesar 0,2, dan *callbacks early stopping*. *Early stopping* diterapkan untuk mencegah terjadinya *overfitting* pada model, yaitu proses pelatihan akan dihentikan jika nilai *loss* pada data validasi mulai meningkat. Peningkatan nilai *loss* pada tahap pelatihan menandakan bahwa model yang dilatih mulai *overfitting* terhadap data pelatihan [19].

Seluruh proses pelatihan model dilakukan dengan menggunakan *hyperparameter* yang telah disebutkan

sebelumnya. Perbedaan utama antara berbagai model yang diuji terletak pada jumlah filter awal dan kedalaman model. Variasi jumlah filter awal dan kedalaman model yang digunakan dalam penelitian ini ditunjukkan pada Tabel III. Dengan menguji berbagai kombinasi ini, diharapkan dapat ditemukan konfigurasi model yang paling optimal untuk tugas segmentasi citra medis dalam PAIP 2023 Challenge.

D. PENGUJIAN MODEL

Pengujian model dilakukan setelah semua model dibuat sesuai dengan spesifikasi yang tercantum dalam Tabel III. Tahapan pengujian ini sangat penting untuk mengevaluasi kinerja model untuk segmentasi semantik. Setelah model selesai dibuat, langkah selanjutnya adalah menguji model tersebut dengan menggunakan data pengujian yang telah disiapkan. Model akan digunakan untuk memprediksi citra masukan dan hasil prediksi ini kemudian dibandingkan dengan *ground truth*. Diharapkan model dapat memprediksi citra sesuai dengan *ground truth*.

Perbandingan antara hasil prediksi dan *ground truth* dilakukan dengan menggunakan indeks Jaccard, yang merupakan metrik yang umum digunakan untuk menghitung akurasi dalam segmentasi semantik [20]. Indeks Jaccard mengukur kesamaan antara dua citra, dalam hal ini antara bagian yang diprediksi oleh model dan bagian yang sebenarnya sesuai *ground truth*. Rumus untuk menghitung indeks Jaccard ditunjukkan oleh (1). Nilai indeks Jaccard berkisar antara 0 hingga 1. Nilai 1 berarti model dapat memprediksi secara sempurna sesuai dengan *ground truth*, sedangkan nilai 0 berarti model sama sekali tidak dapat memprediksi sesuai dengan *ground truth*. Secara teknis, perhitungan indeks Jaccard dilakukan dengan menghitung perbandingan luas irisan dengan luas gabungan dari citra prediksi dengan citra *ground truth*.

$$J(A, B) = \frac{(A \cap B)}{(A \cup B)} = \frac{TP}{TP + FN + FP}. \quad (1)$$

Perhitungan akurasi segmentasi semantik menggunakan indeks Jaccard dapat dilakukan dengan bantuan *confusion matrix*. Dalam konteks segmentasi semantik, hasil prediksi model berupa citra biner, dengan nilai 0 mewakili *background* dan nilai 1 mewakili *region of interest* (ROI). Citra *ground truth* juga merupakan citra biner dengan masing-masing nilai 0 dan 1 sama seperti hasil prediksi. Dengan demikian, perbandingan antara hasil prediksi dan *ground truth* dapat dilakukan dengan mengalkulasi *confusion matrix*. *Confusion matrix* ini memungkinkan klasifikasi dari setiap piksel citra ke dalam empat kategori: *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN). Pengelompokan nilai *confusion matrix* untuk kasus segmentasi semantik diperlihatkan pada Tabel IV.

Proses pengujian dilakukan dengan pemilihan secara acak 57 citra dengan dimensi 256×256 piksel dari data pengujian. Setiap citra ini kemudian diprediksi oleh model dan hasil prediksinya dibandingkan dengan *ground truth*. Setelah semua 57 citra diprediksi, nilai indeks Jaccard dihitung untuk masing-masing citra dan rata-rata dari nilai-nilai tersebut menjadi nilai akhir akurasi dari model yang diuji. Rata-rata ini memberikan gambaran umum tentang kemampuan model dalam melakukan segmentasi pada data pengujian.

Selain akurasi, ada beberapa atribut lain yang dinilai dalam pengujian model, yaitu jumlah parameter, ukuran *file* model, dan waktu eksekusi untuk melakukan prediksi pada 57 citra. Jumlah parameter dan ukuran *file* merupakan bagian dari properti model yang memberikan informasi tentang

TABEL IV
KLASIFIKASI NILAI *CONFUSION MATRIX* PADA SEGMENTASI SEMANTIK

		Nilai <i>Ground Truth</i>	
		<i>Positif (1)</i>	<i>Negatif (0)</i>
Nilai Prediksi	<i>Positif (1)</i>	<i>True Positive</i>	<i>False Positive</i>
	<i>Negatif (0)</i>	<i>False Negative</i>	<i>True Negative</i>

kompleksitas dan ukuran fisik model. Sementara itu, waktu eksekusi mengukur efisiensi model dalam melakukan prediksi, yang menjadi faktor penting ketika model digunakan dalam aplikasi nyata.

Pengujian waktu eksekusi dilakukan dengan menggunakan *central processing unit* (CPU) yang disediakan oleh Kaggle. Penggunaan CPU pada tahapan pengujian dilakukan untuk memastikan bahwa model tidak hanya akurat, tetapi juga efisien dalam hal waktu pemrosesan. Mekanisme pengujian waktu eksekusi model dimulai dari pemuatan model yang sudah tersimpan hingga memprediksi 57 citra medis, sehingga perhitungan waktu dimulai ketika pemuatan model dan selesai ketika selesai memprediksi pada citra ke-57.

Dengan menggabungkan penilaian akurasi, jumlah parameter, ukuran *file*, dan waktu eksekusi, dapat diperoleh gambaran menyeluruh tentang kinerja setiap model. Gambaran menyeluruh ini kemudian digunakan untuk menentukan model terbaik untuk masalah segmentasi semantik sebuah citra medis kanker pankreas.

Perhitungan akurasi pada penelitian ini hanya sebatas dengan menggunakan indeks Jaccard. Beberapa metrik akurasi lain, seperti koefisien Dice, tidak dilakukan karena keterbatasan sumber daya dalam melakukan pengujianya.

III. HASIL PENELITIAN

Model yang telah dibuat dengan beberapa variasi, yaitu jumlah filter awal dan kedalaman, diuji untuk menilai kinerja berdasarkan beberapa metrik pengujian. Metrik ini meliputi akurasi prediksi model yang akan dibandingkan dengan *ground truth*. Selain metrik akurasi yang diukur menggunakan indeks Jaccard, terdapat metrik pengujian lain, yaitu jumlah parameter yang digunakan oleh model, ukuran model, dan waktu eksekusi yang diperlukan untuk memprediksi 57 citra.

Hasil pengujian yang tercantum pada Tabel V menunjukkan bahwa beberapa bagian tidak terisi, yang disebabkan oleh keterbatasan sumber daya komputasi yang tersedia. Keterbatasan ini terutama terlihat pada arsitektur U-Net dengan kedalaman 6, yaitu tidak ada hasil yang dapat dihasilkan untuk semua variasi jumlah filter. Selain itu, pada U-Net dengan jumlah filter awal 128, model tidak dapat dibangun untuk semua tingkat kedalaman yang diinginkan. Sebaliknya, pada arsitektur ELU-Net, semua kombinasi jumlah filter dan kedalaman berhasil dibuat dan diuji, yang menunjukkan bahwa model ini lebih fleksibel dan dapat diterapkan dengan sumber daya yang lebih terbatas dibandingkan dengan U-Net.

Model yang diberi warna biru adalah model yang disarankan oleh masing-masing peneliti. Dari pengujian yang dilakukan, dapat dipahami bahwa model ELU-Net lebih baik secara akurasi, jumlah parameter, waktu eksekusi, dan ukuran model daripada U-Net. Penurunan jumlah parameter hingga 97%, waktu eksekusi 11,6 s, dan ukuran *file* 372,61 MB tidak membuat model ELU-Net tidak lebih akurat dibandingkan dengan U-Net.

TABEL V
 HASIL PENGUJIAN MODEL

Jenis Arsitektur	Jumlah Filter	Jumlah Kedalaman	Jumlah Parameter	Akurasi (Indeks Jaccard)	Waktu Eksekusi (s)	Ukuran Model (MB)	
U-Net	32	4	1.926.149	0,903	4,3	23,26	
		5	7.760.645	0,928	5,3	93,31	
		6	-	-	-	-	
	64	4	4	7.698.437	0,889	12,3	92,53
			5	31.032.837	0,829	15,6	372,61
			6	-	-	-	-
		128	4	-	-	-	-
			5	-	-	-	-
			6	-	-	-	-
	ELU-Net	4	4	52.053	0,867	2,9	0,916
			5	163.109	0,851	3,9	2,25
			6	580.337	0,824	5,2	7,47
8		4	4	205.769	0,854	3,4	2,76
			5	644.697	0,897	4,0	8,1
			6	2.313.537	0,830	6,3	28,26
		16	4	818.193	0,812	5,2	10,1
			5	2.570.417	0,867	5,5	31,2
			6	9.238.529	0,857	9,1	111,34

Secara keseluruhan, model U-Net mendapatkan akurasi lebih tinggi daripada ELU-Net. Model terbaik U-Net memiliki akurasi 0,928 atau 92,8% pada model dengan jumlah filter 32 dan kedalaman 5. Sementara itu, ELU-Net hanya 0,897 atau 89,7% dengan jumlah filter 8 dan kedalaman 5. Selisih akurasi tersebut didapatkan U-Net dengan menggunakan sumber daya yang besar, seperti jumlah parameter, yang membuat waktu eksekusi dan ukuran model lebih besar juga. Model ELU-Net yang memiliki selisih akurasi 3,1% dapat menurunkan jumlah parameter hingga 91,6%, yang membuat penurunan waktu eksekusi hingga 1,3 s dan ukuran model 85,21 MB. Penurunan jumlah parameter, waktu eksekusi, dan ukuran model yang cukup signifikan ini merupakan keuntungan. Akan tetapi, di sisi lain, terdapat kelemahan yaitu akurasi yang juga berkurang. Akurasi yang berkurang merupakan *trade-off* yang cukup logis karena tujuan dari pengembangan ELU-Net adalah membuat model U-Net yang lebih efisien.

Jika membahas tentang akurasi saja, model U-Net lebih unggul daripada ELU-Net. Keunggulan ini tercermin dalam kemampuan U-Net untuk menghasilkan prediksi yang lebih akurat pada tugas segmentasi citra medis. Namun, keunggulan dalam akurasi ini memiliki konsekuensi yang cukup signifikan terhadap penggunaan sumber daya, terutama untuk komputasi. U-Net membutuhkan waktu eksekusi yang lebih lama dan memiliki ukuran model yang lebih besar dibandingkan dengan ELU-Net. Hal ini disebabkan oleh arsitektur U-Net yang kompleks, yang ditandai dengan lebih banyak parameter yang berfungsi untuk menangkap detail gambar secara lebih menyeluruh.

Di sisi lain, ELU-Net menawarkan pendekatan yang lebih efisien dalam hal waktu eksekusi dan ukuran model, dengan menggunakan pendekatan berupa pengembangan dari *skip connection* U-Net, yaitu *deep skip connection*. Meskipun ELU-Net menghasilkan akurasi yang sedikit lebih rendah dibandingkan U-Net, perbedaannya tidak terlalu signifikan. Selisih akurasi model terbaik dari masing-masing arsitektur, sebagaimana tercatat dalam Tabel V, hanya sekitar 3,1%.

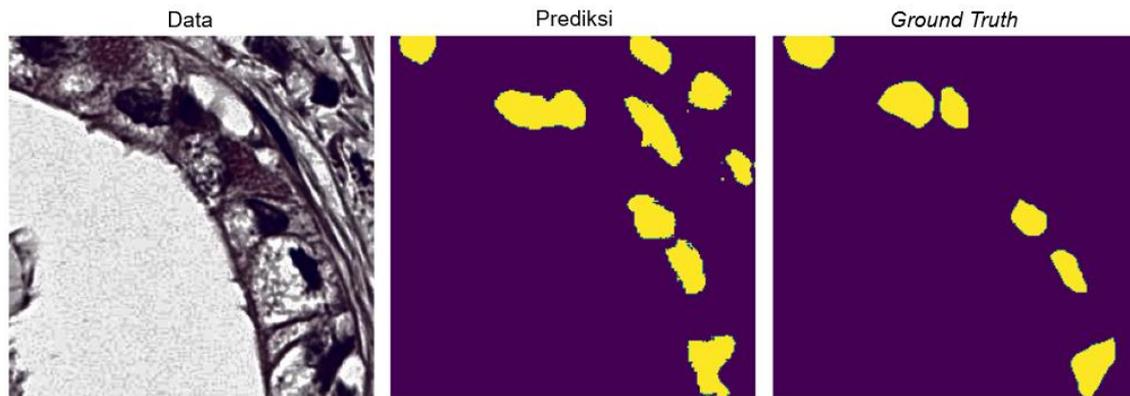
Untuk memvalidasi hasil tersebut, diperlukan pengamatan kasat mata terhadap prediksi yang dihasilkan oleh kedua model. Pengamatan ini dilakukan dengan membandingkan hasil

prediksi dari model dengan *ground truth*. Walaupun pengamatan kasat mata ini bukan merupakan metode pengujian utama, langkah ini penting untuk memastikan bahwa model ELU-Net memberikan hasil yang mendekati akurasi U-Net.

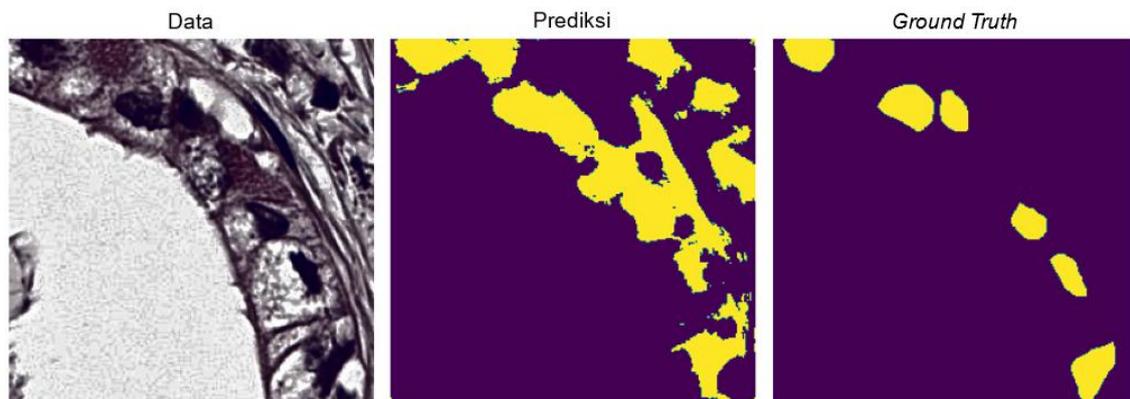
Dari hasil prediksi menggunakan model ELU-Net, didapatkan hasil segmentasi yang sangat mendekati hasil sebenarnya, seperti diperlihatkan dalam Gambar 5. Hal ini dapat dilihat dengan membandingkan hasil prediksi dengan *ground truth* yang tersedia. Demikian juga, pada Gambar 6, prediksi dengan model U-Net menghasilkan citra yang mendekati *ground truth*. Namun, terdapat perbedaan dalam cara kedua model menangani area tertentu. Misalnya, ELU-Net mampu melakukan prediksi yang baik pada beberapa bagian citra, tetapi masih keliru dengan ada sedikit area yang diidentifikasi sebagai sel kanker. Di sisi lain, U-Net juga membuat prediksi yang baik, tetapi dalam beberapa kasus, model ini melakukan *over-segmentation*, yakni memprediksi area yang seharusnya tidak diklasifikasikan sebagai sel kanker.

Dari hasil pengamatan ini, dapat disimpulkan bahwa perbedaan hasil prediksi dari kedua model tidak terlalu signifikan ketika dilihat secara kasat mata. Meskipun U-Net unggul dari segi akurasi berdasarkan nilai numerik, ELU-Net tetap dapat menghasilkan prediksi yang cukup mendekati, dengan keuntungan dalam hal efisiensi komputasi. Efisiensi komputasi ini bukanlah hal yang dapat diabaikan.

Perbedaan hasil uji dari kedua model disebabkan oleh sebuah faktor utama, yaitu adanya pengembangan *skip connection*. Arsitektur ELU-Net merupakan pengembangan dari U-Net yang berfokus untuk mengembangkan *skip connection* dari model U-Net. Beberapa penelitian sebelumnya, seperti U-Net++, merupakan pengembangan dari U-Net yang juga memiliki perbedaan dari sisi *skip connection*. U-Net++ mengembangkan beberapa jenis model dengan berbagai kedalaman dan hasilnya menunjukkan bahwa pengaruh *skip connection* membuat model yang tidak terlalu dalam dapat menyamai akurasi dengan model yang lebih dalam. Penelitian sebelumnya melakukan perubahan dengan menambahkan dan membuang beberapa *skip connection* seperti yang dilakukan dengan arsitektur U-Net++ [12].



Gambar 5. Hasil prediksi dengan ELU-Net.



Gambar 6. Hasil prediksi dengan U-Net.

IV. KESIMPULAN

Arsitektur U-Net dan ELU-Net menghasilkan akurasi yang cukup baik dalam melakukan segmentasi semantik. Model terbaik dari keduanya menghasilkan akurasi di atas 89%. Dari model terbaik pada masing-masing arsitektur diketahui bahwa U-Net memiliki akurasi yang lebih tinggi daripada ELU-Net. Namun, akurasi yang tinggi ini memerlukan sumber daya yang lebih banyak pula. Di sisi lain, ELU-Net dapat menghasilkan model dengan penggunaan sumber daya yang lebih rendah, tetapi akurasinya sedikit lebih kecil daripada U-Net. Selisih akurasi ini merupakan hal yang wajar mengingat tujuan utama ELU-Net adalah membuat pengembangan model U-Net yang lebih sedikit sumber daya dengan hasil yang cukup akurat. Hal ini dapat dibuktikan dari akurasi model terbaik ELU-Net, yang mencapai 89,7%.

Data yang digunakan pada penelitian ini tergolong sedikit untuk membuat sebuah model *deep learning*, sehingga perlu pengembangan lebih lanjut terutama dalam mengumpulkan data penelitian berupa citra H&E kanker pankreas. Selain itu, dapat juga dilakukan pengembangan metode lain dalam melakukan augmentasi data. Namun, dalam pandangan lain, data yang terbatas bisa menjadi sebuah inovasi dalam pengembangan arsitektur *deep learning* dengan minimnya data.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak terdapat konflik kepentingan.

KONTRIBUSI PENULIS

Konseptualisasi, Algi Fari Ramdhani, Yudi Widhiyasa, dan Setiadi Rachmat; metodologi, Algi Fari Ramdhani, Yudi Widhiyasa, dan Setiadi Rachmat; perangkat lunak, Algi Fari Ramdhani; validasi, Yudi Widhiyasa dan Setiadi Rachmat;

analisis formal, Yudi Widhiyasa; investigasi, Algi Fari Ramdhani dan Yudi Widhiyasa; sumber daya, Yudi Widhiyasa dan Setiadi Rachmat; kurasi data, Yudi Widhiyasa; penulisan—penyusunan draf asli, Algi Fari Ramdhani; penulisan—peninjauan dan penyuntingan, Yudi Widhiyasa dan Setiadi Rachmat; visualisasi, Yudi Widhiyasa; pengawasan, Yudi Widhiyasa dan Setiadi Rachmat; administrasi proyek, Yudi Widhiyasa dan Setiadi Rachmat; akuisisi pendanaan, Algi Fari Ramdhani;

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Politeknik Negeri Bandung atas bantuan pendanaan dalam penelitian ini. Terima kasih juga diberikan kepada penyelenggara PAIP 2023 Challenge atas data penelitian yang diberikan dan pesan “*De-identified pathology images and annotations used in this research were prepared and provided by the Seoul National University Hospital by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: H118C0316).*”

REFERENSI

- [1] World Health Organization (WHO), “Cancer.” Tanggal akses: 16-Jun-2024. [Online]. Tersedia: <https://www.who.int/news-room/fact-sheets/detail/cancer>
- [2] J. Cai dkk., “Advances in the epidemiology of pancreatic cancer: Trends, risk factors, screening, and prognosis”, *Cancer Lett.*, vol. 520, hal. 1–11, Nov. 2021, doi: 10.1016/j.canlet.2021.06.027.
- [3] J.S. de Moor dkk., “Cancer survivors in the United States: Prevalence across the survivorship trajectory and implications for care,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 22, no. 4, hal. 561–570, Apr. 2013, doi: 10.1158/1055-9965.EPI-12-1356.
- [4] S. Bharati, P. Podder, dan M.R.H. Mondal, “Artificial neural network based breast cancer screening: A comprehensive review,” *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 12, hal. 125–137, 2020.

- [5] D. Veiga-Canuto dkk., "Comparative multicentric evaluation of inter-observer variability in manual and automatic segmentation of neuroblastic tumors in magnetic resonance images," *Cancers*, vol. 14, no. 15, hal. 1-15, Agu. 2022, doi: 10.3390/cancers14153648.
- [6] M.H. Aziz dan A.A. Abdulla, "Computer-aided diagnosis for the early breast cancer detection," *UHD J. Sci. Technol.*, vol. 7, no. 1, hal. 7-14, Jan. 2023, doi: 10.21928/uhdjst.v7n1y2023.pp7-14.
- [7] A. Rakhlin dkk., "Breast tumor cellularity assessment using deep neural networks," dalam *2019 IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, 2019, hal. 371-380, doi: 10.1109/ICCVW.2019.00048.
- [8] M. Torres-Velazquez, W.-J. Chen, X. Li, dan A.B. McMillan, "Application and construction of deep learning networks in medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, hal. 137-159, Mar. 2021, doi: 10.1109/TRPMS.2020.3030611.
- [9] E. Sudarshan dkk., "Deep learning for the detection and classification of brain tumors using CNN," *AIP Conf. Proc.*, vol. 2971, no. 1, Jun. 2024, Art. No. 020031, doi: 10.1063/5.0196072.
- [10] O. Ronneberger, P. Fischer, dan T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", dalam *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. Wells., A. Frangi, Eds., Cham, Swiss: Springer, 2015, hal. 234-241, doi: 10.1007/978-3-319-24574-4_28.
- [11] S. Arvind, J.V. Tembhurne, T. Diwan, dan P. Sahare, "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest x-rays," *Results Eng.*, vol. 17, hal. 1-9, Mar. 2023, doi: 10.1016/j.rineng.2023.100929.
- [12] Y. Deng, Y. Hou, J. Yan, dan D. Zeng, "ELU-Net: An efficient and lightweight U-Net for medical image segmentation," *IEEE Access*, vol. 10, hal. 35932-35941, Mar. 2022, doi: 10.1109/ACCESS.2022.3163711.
- [13] K. Manasa dan G.V. Murthy, "Skin Cancer Detection Using VGG-16," *Eur. J. Mol. Clin. Med.*, vol. 8, no. 1, hal. 1419-1426, Jan. 2021.
- [14] S. Jha dkk., "Neutrosophic image segmentation with Dice Coefficients," *Measurement*, vol. 134, hal. 762-772, Feb. 2019, doi: 10.1016/j.measurement.2018.11.006.
- [15] R. Awan dkk., "Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images," *Sci. Rep.*, vol. 7, hal. 1-12, Des. 2017, doi: 10.1038/s41598-017-16516-w.
- [16] S. Roy, K. Bhalla, S. Patel, dan R. Patel, "Statistical analysis of histogram equalization techniques for medical image enhancement: A brief study," *SSRN Electronic Journal*, Mar. 2022, doi: 10.2139/ssrn.4049614.
- [17] X. Zhang dkk., "FlipCAM: A feature-level flipping augmentation method for weakly supervised building extraction from high-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, hal. 1-17, Jan. 2024, doi: 10.1109/TGRS.2024.3360276.
- [18] I. Saputra dan D.A. Kristiyanti, *Machine Learning untuk Pemula*. Bandung, Indonesia: Informatika, 2022.
- [19] H. Li dkk., "Keeping deep learning models in check: A history-based approach to mitigate overfitting," *IEEE Access*, vol. 12, hal. 70676-70689, Mei 2024, doi: 10.1109/ACCESS.2024.3402543.
- [20] D. Ogwok dan E.M. Ehlers, "Jaccard index in ensemble image segmentation: An approach," dalam *Proc. 2022 5th Int. Conf. Comput. Intelli. Intell. Syst. (CHIS '22)*, 2022, hal. 9-14, doi: 10.1145/3581792.3581794.