

MULTITASK LEARNING UNTUK ANALISIS LUKA BAKAR MENGGUNAKAN U-NET, EFFICIENTNETB4, DAN VISION TRANSFORMER

Ayu Lutfiah^{1*}, Nirmala², Jeni Fajarwati³, Adha Mashur Sajiah⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo; Jl. H.E.A. Mokodompit, Kampus Hijau Bumi Tridharma, Anduonohu, Kec. Kambu, Kota Kendari, Sulawesi Tenggara 93232

Keywords: *Deep Learning; Burn Wound; Segmentation; MultiTask Learning; Vision Transformer.*

Correspondent Email:
nirmala.23012@gmail.com

Abstrak. Luka bakar merupakan cedera serius yang dapat mengancam jiwa dan memerlukan penanganan cepat sesuai tingkat keparahannya. Keterbatasan tenaga medis terlatih di daerah terpencil mendorong pengembangan sistem diagnosis otomatis berbasis kecerdasan buatan. Penelitian ini mengembangkan sistem analisis luka bakar berbasis *deep learning* dengan pendekatan *MultiTask Learning* (MTL) yang mampu melakukan segmentasi area luka, klasifikasi derajat keparahan (I, II, dan III), serta estimasi rasio area luka secara bersamaan dalam satu model. Tiga arsitektur dibandingkan, yaitu U-Net, EfficientNetB4, dan Vision Transformer (ViT), menggunakan 1.846 citra luka bakar dari Roboflow dan Kaggle. *Dataset* dibagi menjadi data latih, validasi, dan uji dengan rasio 80:10:10 menggunakan *stratified sampling*. Hasil menunjukkan bahwa U-Net memberikan performa segmentasi terbaik dengan nilai Dice 0,6773 dan IoU 0,5121. EfficientNetB4 mencapai akurasi klasifikasi tertinggi sebesar 77,30%, sedangkan Vision Transformer (ViT) unggul pada estimasi rasio area luka dengan MAE 0,0819 dan R^2 0,5107. Temuan ini menunjukkan bahwa tidak ada satu arsitektur yang unggul pada semua tugas, sehingga pemilihan model perlu disesuaikan dengan prioritas klinis. Sistem ini berpotensi mendukung diagnosis luka bakar di fasilitas kesehatan dengan sumber daya terbatas.



Copyright © 2026 The Author(s),
Published by Jurnal Informatika dan
Teknik Elektro Terapan. This is an
open-access article distributed under
the terms of the Creative Commons
Attribution-NonCommercial 4.0
International License (CC BY-NC
4.0).

Abstract. *Burn injuries are serious and potentially life-threatening conditions that require prompt treatment according to their severity. Limited access to trained medical personnel in remote areas has encouraged the development of automated artificial intelligence-based diagnostic systems. This study develops a burn analysis system based on deep learning using a MultiTask Learning (MTL) approach that simultaneously performs wound segmentation, severity classification (Degrees I, II, and III), and wound area ratio estimation within a single model. Three architectures, U-Net, EfficientNetB4, and Vision Transformer (ViT), were compared using 1,846 burn images collected from the Roboflow and Kaggle platforms. The dataset was divided into training, validation, and testing sets at an 80:10:10 ratio using stratified sampling. Results show that U-Net achieved the best segmentation performance with a Dice score of 0,6773 and an IoU of 0,5121. EfficientNetB4 obtained the highest classification accuracy of 77.30%, while Vision Transformer (ViT) achieved the best wound area ratio estimation with an MAE of 0,0819 and an R^2 of 0,5107. These findings indicate that no single architecture is superior across all tasks; therefore, model selection should be tailored to specific clinical priorities. The proposed system has the potential to support burn diagnosis in resource-limited healthcare facilities.*

1. PENDAHULUAN

Luka bakar merupakan salah satu kondisi medis gawat darurat yang membutuhkan penilaian cepat dan akurat. Menurut Niri dkk., penilaian yang tepat terhadap luas serta tingkat keparahan luka bakar sangat menentukan keputusan klinis, mulai dari estimasi kebutuhan cairan, pemilihan terapi, hingga perencanaan tindakan bedah [1]. Secara global, luka bakar menyebabkan sekitar 180.000 kematian per tahun, dengan mayoritas terjadi di negara berpenghasilan rendah dan menengah yang memiliki keterbatasan akses terhadap tenaga medis terlatih [2]. Cassidy dkk. menyatakan bahwa metode konvensional seperti Aturan Sembilan (*Rule of Nines*) dan Diagram Lund-Browder selama ini menjadi acuan utama, namun evaluasi secara manual bersifat subjektif dan sangat bergantung pada pengalaman klinis tenaga medis, sehingga rentan terhadap ketidakkonsistenan antarpenilai [3].

Kemajuan pesat dalam bidang *deep learning* telah membuka peluang bagi solusi analisis citra medis yang lebih objektif dan terstandarisasi. Niri dkk. mengembangkan sistem segmentasi luka bakar menggunakan empat model *deep learning* (U-Net, PSPNet, DeepLabV3, Mask R-CNN) dengan DeepLabV3+ mencapai presisi tertinggi sekitar 0,90, namun penelitian ini tidak menyertakan klasifikasi derajat keparahan dan hanya berfokus pada satu aspek tugas saja [1]. He dkk. mengusulkan model *multitask* berbasis Swin Transformer dengan pendekatan *semi-supervised* untuk mengatasi keterbatasan data berlabel pada citra medis, meskipun validasinya terbatas pada *dataset* tertutup tanpa perbandingan lintas arsitektur [4]. Li dkk. mengembangkan model *Deep Fusion Network* (Dfusion) berbasis DeepLabV3-ResNet101 dengan mekanisme *channel attention* dan *spatial attention* untuk segmentasi sekaligus klasifikasi keparahan luka bakar pediatrik, mencapai akurasi 97,65% dan *Dice* 87,66%, namun tidak menyertakan estimasi rasio area luka sebagai indikator klinis [5]. Sementara itu, R. Zhang dkk. melakukan survei komprehensif terhadap analisis citra luka berbasis *deep learning* yang mencakup klasifikasi, deteksi, dan segmentasi pada berbagai jenis luka seperti luka bakar, ulkus kaki diabetik, dan ulkus

tekanan, namun tidak mengintegrasikan ketiga tugas tersebut dalam satu kerangka analisis yang koheren [6].

Dari tinjauan terhadap penelitian-penelitian tersebut, terdapat beberapa kesenjangan yang signifikan. Pertama, sebagian besar studi hanya menangani satu tugas, yaitu segmentasi atau klasifikasi saja, tanpa mengintegrasikan keduanya bersama estimasi rasio area luka dalam satu kerangka analisis yang koheren. Kedua, belum ada penelitian yang secara sistematis membandingkan tiga arsitektur mutakhir U-Net (berbasis *encoder-decoder* untuk segmentasi piksel), EfficientNetB4 (CNN efisien berbasis *compound scaling*), dan *Vision Transformer/ViT* (berbasis mekanisme *self-attention* global) dalam domain luka bakar secara bersamaan. Ketiga, tidak adanya perbandingan lintas arsitektur yang komprehensif menyulitkan pemilihan model yang paling sesuai untuk skenario klinis tertentu. Kebaruan penelitian ini terletak pada pendekatan *MultiTask Learning* (MTL) yang mengintegrasikan tiga tugas klinis sekaligus dalam satu kerangka evaluasi komparatif, sehingga menghasilkan panduan pemilihan arsitektur yang berbasis bukti (*evidence-based*) untuk aplikasi klinis nyata.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk: (1) mengembangkan sistem segmentasi area luka bakar pada tingkat piksel menggunakan U-Net, (2) melakukan klasifikasi tingkat keparahan luka bakar ke dalam tiga kategori klinis (Derajat I, II, dan III) menggunakan EfficientNetB4, serta (3) mengestimasi rasio area luka bakar dalam citra menggunakan *Vision Transformer* (ViT). Penelitian ini berupaya menjawab pertanyaan: arsitektur *deep learning* manakah yang paling optimal untuk masing-masing tugas klinis dalam analisis luka bakar, dan bagaimana perbandingan kinerjanya secara komprehensif? Sistem yang diusulkan diharapkan dapat menjadi alat bantu pengambilan keputusan klinis yang lebih objektif, meningkatkan konsistensi diagnosis terutama di fasilitas kesehatan dengan keterbatasan sumber daya manusia.

2. TINJAUAN PUSTAKA

2.1 Luka Bakar dan Klasifikasi Derajat Keparahan

Luka bakar diklasifikasikan berdasarkan kedalaman jaringan yang terdampak ke dalam tiga derajat. Derajat I (*Superficial*) hanya mengenai lapisan epidermis, ditandai kemerahan dan nyeri tanpa pembentukan lepuhan. Derajat II (*Partial Thickness*) menembus hingga lapisan dermis, ditandai lepuhan, luka basah, dan nyeri hebat. Derajat III (*Full Thickness*) merusak seluruh lapisan kulit hingga jaringan di bawahnya, ditandai luka kering berwarna putih atau hitam, dan tidak terasa nyeri karena saraf sudah rusak. Cassidy dkk. menyatakan bahwa metode konvensional seperti Aturan Sembilan (*Rule of Nines*) dan Diagram Lund-Browder yang selama ini menjadi acuan utama bersifat subjektif dan rentan terhadap ketidakkonsistenan antarpemilai [3].

2.2 Segmentasi Citra Medis Berbasis *Deep Learning*

Gao dkk. menjelaskan bahwa segmentasi citra medis berbasis *deep learning* telah berkembang pesat dan mampu menghasilkan pemetaan area yang lebih objektif dibandingkan metode manual [7]. Minaee dkk. menambahkan bahwa Dice Coefficient dan *Intersection over Union* (IoU) merupakan metrik standar untuk mengukur kualitas segmentasi, di mana Dice Loss lebih tahan terhadap ketidakseimbangan kelas dibandingkan *Binary Cross-Entropy* [8]. Chtouki dkk. memperkuat hal ini dengan membuktikan bahwa pendekatan *multi-task deep learning* mampu melakukan segmentasi dan klasifikasi secara simultan pada citra medis [9].

2.3 Arsitektur *Deep Learning* yang Digunakan

2.3.1 U-Net

Menurut Gao dkk., U-Net dibangun dengan arsitektur *encoder-decoder* simetris yang mempertahankan detail spasial melalui *skip connection* dari tiap level *encoder* ke *decoder* [7]. Struktur ini terbukti efektif dalam menjaga informasi tepi dan detail lokasi luka, yang sangat penting dalam segmentasi citra medis.

2.3.2 *EfficientNetB4*

Y. Zhang dkk. menyatakan bahwa arsitektur CNN berbasis *deep learning* mampu mengekstraksi fitur visual secara hierarkis dengan akurasi klasifikasi tinggi [10]. A. Zhang dkk. menjelaskan bahwa pendekatan *transfer learning* menggunakan model *pre-trained* mempercepat konvergensi dan meningkatkan efisiensi pelatihan [11]. Abdelrahman dan Viriri membuktikan bahwa EfficientNetB4 sebagai *encoder* pada arsitektur U-Net menghasilkan IoU tertinggi dibanding varian EfficientNet lainnya untuk segmentasi citra medis [12]. Penelitian sebelumnya di domain klasifikasi citra medis kulit menggunakan arsitektur EfficientNet-B3 dengan pendekatan *transfer learning* dan *fine-tuning* berhasil mencapai akurasi 96,61% dalam mengklasifikasikan lima jenis penyakit kulit wajah, yang menunjukkan efektivitas arsitektur berbasis *compound scaling* dalam mengekstraksi fitur diskriminatif citra medis [13].

2.3.3 *Vision Transformer (ViT)*

Menurut Maurya dan Mirza, arsitektur berbasis *dual-branch* yang menggabungkan CNN dan Transformer terbukti efektif dalam klasifikasi luka, di mana cabang CNN menangkap fitur lokal sementara cabang Transformer mengkodekan informasi kontekstual global pada berbagai resolusi, menghasilkan akurasi rata-rata 91,20% untuk klasifikasi biner dan 89,80% untuk klasifikasi multikelas [14]. Y. Zhang dkk. menegaskan bahwa mekanisme *self-attention* global ini unggul dalam menangkap korelasi makroskopis seluruh citra yang tidak dapat dijangkau oleh konvolusi lokal [10]. Penerapan arsitektur Vision Transformer (ViT) pada domain citra medis terbukti mampu memanfaatkan mekanisme global *self-attention* untuk menangkap pola visual secara menyeluruh, sebagaimana ditunjukkan pada penelitian deteksi stunting berbasis citra tubuh balita yang mencapai akurasi 98% dengan model ViT-base [15].

2.4 *MultiTask Learning (MTL)*

Kendall dkk. merumuskan total *loss* pada MTL sebagai kombinasi berbobot dari *loss* setiap tugas, dengan bobot yang dikalibrasi untuk menyeimbangkan skala gradien dari tugas-tugas yang berbeda [16]. Pendekatan ini

memungkinkan satu *backbone* bersama menghasilkan beberapa prediksi sekaligus, yaitu segmentasi, klasifikasi, dan regresi, secara paralel dalam satu kerangka pelatihan.

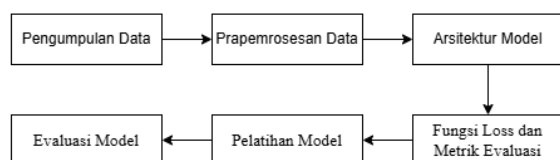
2.5 Fungsi Loss dan Metrik Evaluasi

Minaee dkk. dan Chtouki dkk. menjelaskan bahwa *Binary Cross-Entropy* menangani klasifikasi piksel secara biner, sedangkan Dice Loss mengukur tumpang tindih langsung antara prediksi dan *ground truth* [8],[9]. A. Zhang dkk. menyebutkan bahwa *Mean Squared Error* (MSE) digunakan untuk tugas regresi, sedangkan *Categorical Cross-Entropy* digunakan untuk tugas klasifikasi multikelas [11]. Szeliski menegaskan bahwa Dice Coefficient dan IoU merupakan metrik standar evaluasi segmentasi, sementara MAE, RMSE, dan R^2 lazim digunakan untuk menilai akurasi regresi [17].

3. METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini membandingkan tiga arsitektur *deep learning* U-Net, EfficientNetB4, dan Vision Transformer (ViT) menggunakan pendekatan *MultiTask Learning* (MTL) untuk segmentasi, klasifikasi derajat, dan estimasi luas luka bakar secara bersamaan. Seluruh eksperimen dijalankan di Google Colaboratory dengan dukungan GPU NVIDIA.



Gambar 1. Desain Penelitian

3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dua sumber yang kemudian digabungkan menjadi satu *dataset* terintegrasi. Sumber pertama adalah *dataset* publik *Burn Detection 1* dari platform Roboflow, yang menyediakan 1,704 citra luka bakar beserta anotasi segmentasi berbasis piksel dan label keparahan dalam format COCO. *Dataset* ini sudah terbagi ke dalam tiga subset, yakni 1,197 citra latih, 336 citra validasi, dan 171 citra uji. Sumber kedua berasal dari Kaggle, dengan 142 citra yang dipilih berdasarkan kualitas dan relevansinya. Berbeda dari sumber pertama,

citra dari Kaggle belum memiliki anotasi, sehingga perlu dilakukan pelabelan secara manual menggunakan standar format COCO yang sama mencakup penandaan area luka dan penentuan derajat keparahannya (Derajat I, II, atau III). Setelah kedua sumber digabungkan, total citra yang tersedia berjumlah 1,846 sampel. Seluruh data ini kemudian dibagi ulang dengan rasio 80:10:10, sehingga menghasilkan 1,476 citra untuk pelatihan, 185 citra untuk validasi, dan 185 citra untuk pengujian. Agar proporsi setiap kelas keparahan tetap seimbang di masing-masing subset, proses pembagian data dilakukan dengan teknik *stratified sampling*.

3.3. Prapemrosesan Data

Prapemrosesan dilakukan melalui dua tahap utama:

1. Konversi Anotasi ke Mask

Anotasi COCO dari kedua *dataset* dikonversi menjadi *binary mask* per piksel menggunakan *pycocotools*. Setiap poligon anotasi digambar ke dalam *mask* berukuran sesuai dimensi gambar asli. Label kelas keparahan dipetakan ke tiga derajat (Derajat I, II, III) menggunakan *keyword matching* pada nama kategori, dengan *voting* mayoritas apabila satu gambar memiliki lebih dari satu anotasi. Rasio area luka dihitung sebagai proporsi piksel luka terhadap total piksel gambar, lalu di *clipping* pada rentang 0,001 sampai 0,95 untuk menghindari nilai ekstrem.

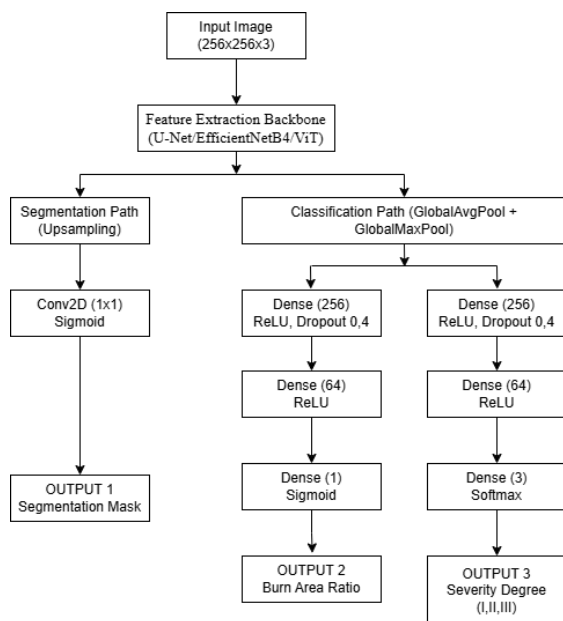
2. Normalisasi dan Augmentasi

Gambar

Seluruh gambar dan *mask* di *resample* ke ukuran 256×256 piksel. Normalisasi menggunakan *mean* dan standar deviasi ImageNet ([0,485, 0,456, 0,406] dan [0,229, 0,224, 0,225]). Pada data latih diterapkan augmentasi meliputi *horizontal* atau *vertical flip*, rotasi acak, *shift-scale-rotate*, *gaussian noise*, serta variasi kecerahan dan saturasi menggunakan library *Albumentations*. Data validasi dan *test* hanya melalui *resize* dan normalisasi tanpa augmentasi.

3.4. Arsitektur Model Deep Learning

Ketiga model dirancang menghasilkan tiga *output* sekaligus *mask* segmentasi area luka, rasio area luka dalam citra, dan klasifikasi derajat keparahan (I,II,III).



Gambar 2. Arsitektur *MultiTask Learning*

Gambar 2. mengilustrasikan arsitektur *MultiTask Learning* yang digunakan dalam penelitian ini. Dari satu *backbone* bersama (U-Net, EfficientNetB4, atau ViT), model bercabang menjadi dua jalur yang berjalan secara paralel, yaitu jalur segmentasi untuk menghasilkan mask area luka dan jalur klasifikasi untuk menghasilkan prediksi derajat keparahan serta rasio area luka secara bersamaan.

3.4.1 U-Net

Encoder U-Net menurunkan resolusi melalui empat tahap konvolusi dengan filter bertingkat dari 64 hingga 512 menggunakan *max-pooling*, dilanjutkan *bottleneck* 1.024 filter. Decoder mengembalikan resolusi melalui *Conv2DTranspose* dengan *skip connection* dari tiap level encoder. Untuk prediksi klasifikasi derajat dan estimasi rasio area luka, fitur dari *bottleneck* digabungkan menggunakan *GlobalAveragePooling* dan *GlobalMaxPooling*, kemudian diteruskan ke lapisan *Dense* 256 dengan *dropout* 0,4.

3.4.2 EfficientNetB4

EfficientNetB4 digunakan sebagai encoder dengan bobot pra-latih ImageNet. Decoder segmentasi dibangun dari lima lapis *Conv2DTranspose* dengan filter bertingkat dari 512 hingga 32. Untuk prediksi klasifikasi derajat dan estimasi rasio area luka, fitur dari *output* encoder digabungkan menggunakan *GlobalAveragePooling* dan *GlobalMaxPooling*, kemudian diteruskan ke lapisan *Dense* 256 dengan *dropout* 0,4.

3.4.3 Vision Transformer (ViT)

Citra dipecah menjadi 256 *patch* berukuran 16×16 piksel dan diproyeksikan ke dimensi 256 dengan *positional embedding*. Enam blok Transformer dengan delapan *attention head* memproses seluruh *patch* secara global. Decoder segmentasi merekonstruksi ruang spasial melalui empat lapis *Conv2DTranspose* dengan filter bertingkat dari 128 hingga 16. Untuk prediksi klasifikasi derajat dan estimasi rasio area luka, fitur dari seluruh token *patch* diringkas menggunakan *GlobalAveragePooling1D* dengan *dropout* 0,3 sebelum diteruskan ke lapisan *output* akhir.

3.5 Fungsi Loss

Total *loss* model diformulasikan sebagai kombinasi berbobot dari tiga *loss* per tugas [16]:

$$L_{\text{total}} = W_{\text{SEG}} \cdot L_{\text{seg}} + W_{\text{AREA}} \cdot L_{\text{area}} + W_{\text{DEG}} \cdot L_{\text{degree}} \quad (1)$$

dengan $W_{\text{SEG}} = 1,0$, $W_{\text{AREA}} = 0,3$, dan $W_{\text{DEG}} = 1,0$. Bobot W_{AREA} ditetapkan lebih kecil untuk mencegah skala *output* regresi mendominasi gradien pelatihan. *Loss* segmentasi (L_{seg}) menggabungkan *Binary Cross-Entropy* dan *Dice Loss*. *Loss* regresi (L_{area}) menggunakan *Mean Squared Error* (MSE), sedangkan *loss* klasifikasi (L_{degree}) menggunakan *Categorical Cross-Entropy*.

3.6 Pelatihan Model

Ketiga model dilatih menggunakan optimizer Adam dengan *learning rate* awal 1×10^{-4} , maksimal 30 *epoch*, dan *batch size* 8. Tiga *callback* diterapkan selama pelatihan: *ModelCheckpoint* menyimpan bobot terbaik berdasarkan *validation loss*; *EarlyStopping* menghentikan pelatihan jika tidak ada peningkatan selama 15 *epoch*, dan

ReduceLROnPlateau menurunkan *learning rate* sebesar faktor 0,5 jika *validation loss* stagnan selama 7 *epoch*, dengan batas minimum 1×10^{-7} .

3.7 Evaluasi Model

Setiap model dievaluasi pada *test set* menggunakan metrik yang disesuaikan per tugas. Tugas segmentasi diukur menggunakan *Dice Coefficient*, *Intersection over Union* (IoU), dan *pixel accuracy*. Tugas regresi rasio area luka diukur menggunakan MAE, RMSE, dan R^2 . Tugas klasifikasi keparahan diukur menggunakan *accuracy*, *precision*, *recall*, dan *F1-Score* per kelas. Selain evaluasi kuantitatif, analisis kualitatif dilakukan dengan membandingkan *mask* prediksi terhadap *ground truth* secara visual untuk menilai kesesuaian bentuk dan batas area luka yang dihasilkan masing-masing arsitektur.

4 HASIL DAN PEMBAHASAN

4.1 Perbandingan Performa Keseluruhan

Ketiga model dievaluasi pada 185 sampel *test set*. Tabel 2 merangkum perbandingan performa ketiganya.

Tabel 2. Perbandingan Performa Model

Metric	U-Net	EfficientNet B4	ViT
Dice	0,6773	0,6294	0,6491
IoU	0,5121	0,4592	0,4805
Pixel Acc	0,8619	0,8561	0,8675
MAE	0,1568	0,1262	0,0819
RMSE	0,1919	0,1716	0,1229
R^2	-0,1924	0,0470	0,5107
Class Acc	76,76%	77,30%	75,14%

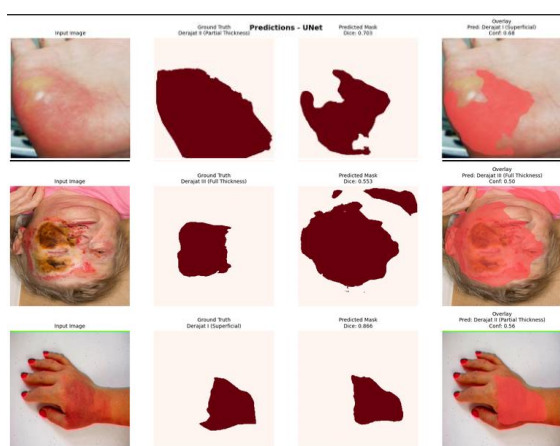
Hasil evaluasi menunjukkan bahwa tidak ada satu arsitektur yang dominan di semua aspek. U-Net unggul pada tugas segmentasi dengan Dice tertinggi 0,6773 dan IoU 0,5121. EfficientNetB4 mencatat *classification accuracy* terbaik sebesar 77,30%. ViT secara signifikan mengungguli dua model lainnya pada estimasi rasio area luka dengan MAE terendah 0,0819 dan R^2 tertinggi 0,5107. Temuan ini secara eksplisit menjawab pertanyaan penelitian bahwa tidak ada satu arsitektur tunggal yang optimal untuk seluruh tugas klinis analisis luka bakar; setiap arsitektur memiliki keunggulan yang berbeda sesuai karakteristik tugasnya.

Fenomena ini memvalidasi konsep *inductive bias* yang diuraikan Gao dkk. dan Minaee dkk., bahwa setiap desain jaringan membawa asumsi struktural berbeda yang menentukan kemampuannya pada jenis tugas tertentu [7][8]. Keunggulan U-Net pada segmentasi sejalan dengan temuan Niri dkk. bahwa struktur *skip connection* superior dalam mempertahankan detail spasial tepi luka [1]. Dominasi ViT pada estimasi rasio area memperkuat analisis Y. Zhang dkk. bahwa mekanisme *global self-attention* mampu menangkap korelasi makroskopis seluruh citra yang tidak dapat dijangkau konvolusi lokal [10]. Tingginya akurasi klasifikasi EfficientNetB4 mencerminkan *feature trade-off* khas MTL, di mana *backbone* berbasis *compound scaling* dengan bobot pra-latih ImageNet lebih efisien dalam mengalokasikan representasi fitur lintas tugas secara simultan.

Jika dibandingkan dengan penelitian Li dkk. yang mencapai Dice 0,8766 menggunakan *Deep Fusion Network* berbasis DeepLabV3-ResNet101 dengan mekanisme *channel* dan *spatial attention* [5], nilai Dice ketiga model dalam penelitian ini masih lebih rendah. Perbedaan ini dapat dijelaskan oleh beberapa faktor: Li dkk. menggunakan dataset khusus luka bakar pediatrik yang lebih homogen secara visual, arsitektur *fusion* ganda yang secara eksplisit dirancang untuk menggabungkan informasi spasial dan saluran, serta tidak menyertakan tugas regresi rasio area yang dalam penelitian ini turut membagi kapasitas representasi *backbone*. Hal ini mengonfirmasi bahwa penambahan tugas dalam kerangka MTL membawa konsekuensi *trade-off* pada performa segmentasi dibandingkan model *single-task* yang terspesialisasi.

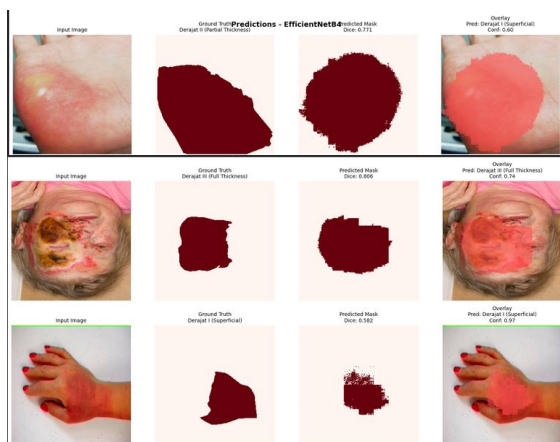
4.2 Hasil Segmentasi Area Luka Bakar

Analisis kualitatif dilakukan dengan membandingkan *mask* prediksi terhadap *ground truth* pada tiga sampel representatif yang mewakili masing-masing derajat keparahan: telapak tangan (Derajat II dengan bula kuning), kepala (Derajat III dengan *eschar* coklat kehitaman), dan tangan (Derajat I dengan eritema merata).



Gambar 3. Prediksi Model U-Net

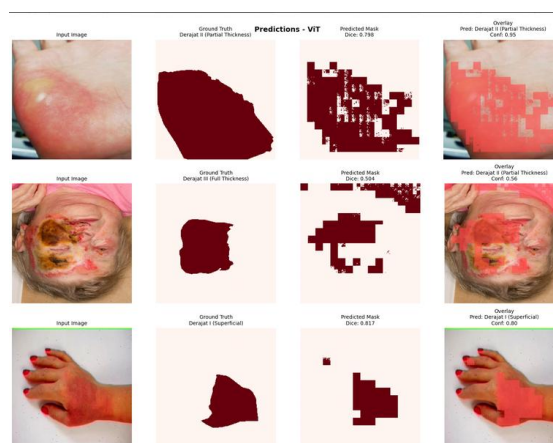
Pada model U-Net, ketiga sampel menunjukkan karakteristik performa yang berbeda. U-Net gagal mendeteksi bula kuning pada sampel telapak tangan sehingga salah mengklasifikasikan sebagai Derajat I (*confidence* 0,68, Dice 0,703), menunjukkan keterbatasan dalam mengenali fitur lokal berukuran kecil yang menjadi pembeda antar derajat. Pada sampel kepala, model benar mengklasifikasikan Derajat III namun dengan *confidence* rendah (0,50) dan Dice hanya 0,553 karena *predicted mask* meluas ke area krusta sekitar *eschar*. Pada sampel tangan, meski Dice tinggi (0,866) karena *mask* presisi mengikuti kontur eritema, model salah mengklasifikasikan sebagai Derajat II (*confidence* 0,56), mengindikasikan bahwa presisi segmentasi tidak selalu berbanding lurus dengan akurasi klasifikasi dalam kerangka MTL.



Gambar 4. Prediksi Model EfficientNetB4

EfficientNetB4 juga gagal mendeteksi bula kuning pada sampel telapak tangan dan salah mengklasifikasikan sebagai Derajat I

(*confidence* 0,60, Dice 0,771) dengan *mask* berbentuk bundar yang kurang mengikuti kontur asli. Pada sampel kepala, kontras warna *eschar* yang tegas terhadap kulit sekitarnya membantu model mengklasifikasikan dengan benar sebagai Derajat III dengan *confidence* tertinggi di antara semua model (0,74) dan Dice terbaik (0,806), mengonfirmasi keunggulan EfficientNetB4 dalam mengekstraksi fitur warna dan tekstur yang kontras. Pada sampel tangan, klasifikasi benar sebagai Derajat I dengan *confidence* sangat tinggi (0,97), namun Dice turun drastis ke 0,582 karena *predicted mask* jauh lebih kecil dari luas eritema aktual, menunjukkan kecenderungan model untuk mengklasifikasikan dengan tepat namun undersegmentasi pada area luka dengan batas yang baur.



Gambar 5. Prediksi Model ViT

ViT menjadi satu-satunya model yang berhasil mengenali pola bula kuning sebagai karakteristik Derajat II, sehingga benar mengklasifikasikan sampel telapak tangan dengan *confidence* tertinggi (0,95) dan Dice terbaik di antara semua model untuk sampel ini (0,798). Kemampuan ini sejalan dengan mekanisme *global self-attention* yang memungkinkan ViT menghubungkan fitur lokal bula dengan konteks keseluruhan citra. Namun pada sampel kepala dengan tekstur paling heterogen (campuran warna coklat, hitam, dan kuning pada *eschar*), ViT justru kesulitan dan salah mengklasifikasikan sebagai Derajat II (*confidence* 0,56) dengan Dice terendah di antara semua model (0,504), karena *predicted mask* hanya menangkap sebagian kecil area secara terputus-putus akibat keterbatasan resolusi *patch* 16×16 dalam merekonstruksi

tekstur halus yang heterogen. Pada sampel tangan, ViT benar mengklasifikasikan sebagai Derajat I dengan *confidence* tinggi (0,80) dan Dice cukup baik (0,817).

Secara keseluruhan, ketiga model menghadapi tantangan serupa dalam mendeteksi fitur pembeda antar derajat yang berukuran kecil atau bertekstur heterogen. Temuan ini konsisten dengan keterbatasan yang diidentifikasi Cassidy dkk. bahwa variabilitas visual tinggi antar sampel luka bakar menjadi hambatan utama sistem otomatis berbasis *deep learning* [3].

4.3 Hasil Regresi Rasio Area Luka

ViT unggul signifikan pada tugas estimasi rasio area luka dengan MAE 0,0819, RMSE 0,1229, dan R^2 0,5107, artinya model mampu menjelaskan sekitar 51% variansi rasio area luka dalam citra. Sebaliknya, U-Net menghasilkan R^2 negatif (-0,1924), menandakan prediksinya lebih buruk dibandingkan sekadar menggunakan nilai rata-rata sebagai prediksi konstan. EfficientNetB4 sedikit lebih baik dengan R^2 0,0470 namun masih sangat rendah dan tidak memiliki daya prediktif yang berarti.

Keunggulan ViT pada tugas ini secara langsung dikaitkan dengan mekanisme *self-attention* global yang memungkinkan model mempertimbangkan distribusi seluruh *patch* citra secara bersamaan untuk memperkirakan proporsi area luka. U-Net dan EfficientNetB4 yang berbasis konvolusi lokal secara struktural kurang mampu mengintegrasikan informasi spasial secara global dalam satu operasi, sehingga estimasi rasio area yang membutuhkan pemahaman konteks keseluruhan citra menjadi kelemahan inherent kedua model tersebut. Temuan ini memperluas bukti empiris analisis Y. Zhang dkk. tentang superioritas ViT dalam tugas yang membutuhkan pemahaman konteks global [10], khususnya dalam domain estimasi kuantitatif citra medis luka bakar yang belum banyak dieksplorasi sebelumnya.

4.4 Hasil Klasifikasi Keparahan Luka Bakar

EfficientNetB4 mencapai *accuracy* tertinggi 77,30%, diikuti U-Net 76,76% dan ViT 75,14%. Perbedaan performa antar kelas secara rinci ditampilkan pada Tabel 3.

Tabel 3. Classification Report Ketiga Model

Metric	U-Net	EfficientNetB4	ViT
Precision D-I	0,71	0,72	0,70
Recall D-I	0,91	0,97	0,79
F1 D-I	0,80	0,83	0,74
Support D-I	66	66	66
Precision D-II	0,90	0,90	0,84
Recall D-II	0,67	0,68	0,70
F1 D-II	0,77	0,78	0,77
Support D-II	107	107	107
Precision D-III	0,48	0,40	0,55
Recall D-III	0,83	0,50	1,00
F1 D-III	0,61	0,44	0,71
Support D-III	12	12	12

Ketiga model menunjukkan pola serupa, performa terbaik pada Derajat II karena jumlah sampelnya paling dominan (107 dari 185 sampel uji), sedangkan Derajat III paling sulit diklasifikasikan akibat ketidakseimbangan data yang ekstrem dengan hanya 12 sampel. Pola ini konsisten dengan fenomena ketidakseimbangan kelas yang umum pada dataset citra medis, sebagaimana diidentifikasi Minaee dkk. bahwa distribusi kelas yang tidak seimbang secara signifikan memengaruhi performa klasifikasi *deep learning* [8].

ViT mencatat *recall* sempurna 1,00 untuk Derajat III dengan F1-Score 0,71, menunjukkan kemampuan mendeteksi seluruh kasus luka bakar paling parah tanpa ada yang terlewat. Dalam konteks klinis, kemampuan ini sangat kritis karena kegagalan mendeteksi Derajat III dapat berakibat fatal bagi pasien. ViT juga mencatat *precision* tertinggi untuk Derajat III (0,55) dibanding U-Net (0,48) dan EfficientNetB4 (0,40), menjadikan ViT sebagai pilihan paling andal untuk deteksi kasus kritis meskipun *accuracy* keseluruhannya tidak tertinggi. Jika dibandingkan dengan akurasi 97,65% yang dicapai Li dkk. [5], perbedaan yang cukup besar ini mencerminkan kompleksitas tambahan dalam penelitian ini berupa penanganan tiga tugas simultan dalam MTL dan keberagaman dataset yang lebih heterogen antara luka bakar anak dan dewasa.

4.5 Implikasi Hasil

Secara teoritis, penelitian ini memberikan bukti empiris bahwa *inductive bias* masing-masing arsitektur *deep learning* secara langsung menentukan keunggulan pada jenis tugas klinis yang berbeda dalam kerangka MTL. Temuan

ini memodifikasi pandangan sebelumnya yang cenderung mencari satu arsitektur terbaik secara universal, dan mengusulkan bahwa pemilihan arsitektur sebaiknya didasarkan pada prioritas tugas klinis yang spesifik.

Secara praktis, hasil penelitian ini mengimplikasikan panduan pemilihan arsitektur berbasis bukti untuk aplikasi klinis: U-Net direkomendasikan untuk aplikasi yang memprioritaskan akurasi segmentasi batas luka, misalnya perencanaan debridemen dan pemantauan penyembuhan; ViT direkomendasikan untuk estimasi luas luka dan deteksi kasus kritis Derajat III yang tidak boleh terlewat; sementara EfficientNetB4 cocok untuk sistem triase awal yang membutuhkan klasifikasi keparahan secara cepat dengan akurasi keseluruhan tertinggi. Lebih jauh, pendekatan *ensemble* yang menggabungkan ketiganya berpotensi mengoptimalkan performa di seluruh tugas secara bersamaan, dan menjadi arah penelitian lanjutan yang menjanjikan terutama di fasilitas kesehatan dengan keterbatasan tenaga medis terlatih.

5 KESIMPULAN

Penelitian ini berhasil mengembangkan dan membandingkan tiga arsitektur *deep learning* berbasis *MultiTask Learning* (MTL) untuk analisis luka bakar secara simultan, mencakup segmentasi area luka, klasifikasi derajat keparahan, dan estimasi rasio area luka. Hasil evaluasi pada 185 sampel *test set* menghasilkan temuan sebagai berikut:

1. U-Net mencapai performa segmentasi terbaik dengan Dice 0,6773 dan IoU 0,5121 berkat struktur *skip connection* yang efektif mempertahankan detail spasial tepi luka, namun menunjukkan kelemahan pada estimasi rasio area luka dengan R^2 negatif (-0,1924).
2. EfficientNetB4 mencapai *classification accuracy* tertinggi 77,30% didukung oleh bobot pra-latih ImageNet yang efisien dalam mengekstraksi fitur diskriminatif, namun menghasilkan F1-Score terendah untuk Derajat III (0,44) yang merupakan kelas paling kritis secara klinis.
3. ViT mengungguli kedua model pada estimasi rasio area luka dengan MAE 0,0819 dan R^2 0,5107 berkat

mekanisme *global self-attention*, serta mencatat *recall* sempurna 1,00 untuk Derajat III, namun memiliki keterbatasan pada rekonstruksi tekstur heterogen akibat resolusi *patch* yang besar.

4. Tidak ada satu arsitektur yang dominan di semua tugas, mengonfirmasi bahwa *inductive bias* masing-masing arsitektur secara langsung menentukan keunggulannya pada jenis tugas yang berbeda. Pemilihan arsitektur sebaiknya disesuaikan dengan prioritas klinis: U-Net untuk segmentasi presisi, ViT untuk estimasi luas luka dan deteksi kasus kritis, serta EfficientNetB4 untuk triase keparahan secara umum.
5. Untuk pengembangan selanjutnya, disarankan memperbesar dataset terutama sampel Derajat III yang masih sangat terbatas (12 sampel), menambahkan anotasi lokasi anatomis untuk memungkinkan estimasi *Total Body Surface Area* (TBSA) secara medis, serta mengeksplorasi pendekatan *ensemble* dari ketiga arsitektur untuk mengoptimalkan performa di seluruh tugas secara bersamaan.

DAFTAR PUSTAKA

- [1] R. Niri dkk., "Wound Segmentation with U-Net Using a Dual Attention Mechanism and Transfer Learning," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 5, pp. 3351–3365, Oct. 2025.
- [2] World Health Organization, "Burns," WHO Fact Sheet, Oct. 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/burns>.
- [3] B. Cassidy, C. Kendrick, N. D. Reeves, J. M. Pappachan, and M. H. Yap, "Deep learning in chronic wound segmentation: a comprehensive review and meta-analysis," *The Visual Computer*, vol. 41, pp. 11885–11908, 2025, doi: 10.1007/s00371-025-04133-y.
- [4] H. He, C. Sun, Z. Wang, and Y. Dan, "Linear Transformer Based U-Shaped Lightweight Segmentation Network," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 29, no. 6, pp. 1319–1328, 2025.
- [5] X. Li, Z. Liu, and L. Liu, "Pediatric BurnNet: Robust Multi-Class Segmentation and Severity

- Recognition under Real-World Imaging Conditions," *SAGE Open Medicine*, vol. 13, pp. 1–12, 2025, doi: 10.1177/20503121251360090.
- [6] R. Zhang, D. Tian, D. Xu, W. Qian, and Y. Yao, "A Survey of Wound Image Analysis Using Deep Learning: Classification, Detection, and Segmentation," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3194529.
- [7] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods," *Tomography*, vol. 11, no. 5, May 2025, doi: 10.3390/tomography11050052.
- [8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.
- [9] K. Chtouki dkk., "Multi-Task Deep Learning for Simultaneous Classification and Segmentation of Cancer Pathologies in Diverse Medical Imaging Modalities," *Onco*, vol. 5, no. 3, Jul. 2025, doi: 10.3390/onco5030034.
- [10] Y. Zhang, J. Wang, J. M. Gorriz, and S. Wang, "Deep Learning and Vision Transformer for Medical Image Analysis," *Journal of Imaging*, vol. 9, no. 7, Jul. 2023, doi: 10.3390/jimaging9070147.
- [11] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*. Cambridge, UK: Cambridge University Press, 2023.
- [12] A. Abdelrahman and S. Viriri, "EfficientNet Family U-Net Models for Deep Learning Semantic Segmentation of Kidney Tumors on CT Images," *Frontiers in Computer Science*, vol. 5, 2023, doi: 10.3389/fcomp.2023.1235622.
- [13] R. A. B. Kusuma, B. Irawan, dan A. Khamid, "Klasifikasi Penyakit Kulit Wajah Menggunakan Convolutional Neural Network EfficientNet-B3," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 14, no. 1, 2026, doi: 10.23960/jitet.v14i1.8721.
- [14] L. Maurya and S. Mirza, "MCTFWC: a multiscale CNN-transformer fusion-based model for wound image classification," *Signal, Image and Video Processing*, May 2025, doi: 10.1007/s11760-025-04072-0.
- [15] R. R. Ar, Agusriyati, dan S. Moka, "Pendeteksian Dini Stunting pada Balita Menggunakan Vision Transformer (ViT) Berbasis Citra Tubuh," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 3S1, 2025, doi: 10.23960/jitet.v13i3S1.7888.
- [16] A. Kendall, Y. Gal, and R. Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, Jun. 2018, pp. 7482–7491.
- [17] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Cham, Switzerland: Springer Nature, 2022.