ANALISIS KOMPARATIF *VISION TRANSFORMER* (VIT) DAN RESNET34 UNTUK KLASIFIKASI DIFERENSIAL SUBTIPE KANKER PARU-PARU BERBASIS CITRA CT SCAN

Rahma Cindy Syuherman , Nasywa Aldira, Farah Dilla, Rahayu A.G. , Syifa Ul Qalbi , Vemilia Zahira, Sri Oktamuliani*

Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas, Limau manis, Padang, 25163, Indonesia
*email: srioktamuliani@sci.unand.ac.id

ABSTRAK

Klasifikasi akurat sub-tipe kanker paru-paru merupakan langkah krusial dalam penentuan terapi dan diagnosis dini. Namun, metode diagnostik konvensional berbasis pemeriksaan manual masih menghadapi tantangan seperti variabilitas antar-pengamat dan keterbatasan waktu. Penelitian ini membandingkan performa dua arsitektur deep learning, Vision Transformer (ViT) dan ResNet34, untuk klasifikasi empat sub-tipe kanker paru berdasarkan citra CT scan menggunakan platform Roboflow. Dataset terdiri atas 288 citra CT yang dibagi menjadi empat kelas: adenokarsinoma, karsinoma sel besar, karsinoma sel skuamosa, dan sel normal. Seluruh citra mengalami pra-pemrosesan standar dan augmentasi berbasis arsitektur. Hasil evaluasi menunjukkan bahwa ResNet34 mencapai performa paling optimal dengan akurasi 95,7% dan F1-score 96,3%, serta stabilitas konvergensi yang lebih baik. ViT tetap menunjukkan kompetitif dengan akurasi 94,3% dan F1-score 93,6%, meskipun lebih sensitif terhadap keterbatasan jumlah data. Analisis kualitatif mengungkapkan bahwa kesalahan klasifikasi terutama terjadi pada kasus ambigu dengan kemiripan morfologis, khususnya antara adenokarsinoma dan karsinoma sel skuamosa. Dalam kondisi tersebut, ViT lebih sering salah mengklasifikasikan, sementara ResNet34 meskipun tepat dalam prediksi, menghasilkan tingkat keyakinan rendah (0,56), yang mengindikasikan keraguan model terhadap sampel sulit. Penelitian ini juga menunjukkan bahwa arsitektur CNN konvensional seperti ResNet34 lebih unggul dibandingkan arsitektur transformer dalam keterbatasan data. Model yang dikembangkan memiliki potensi untuk diintegrasikan dalam sistem pendukung keputusan klinis berbasis AI guna mempercepat proses analisis diagnostik

Kata Kunci: *Kanker Paru; ViT; ResNet34; Deep Learning; CT Scan*

ABSTRACT

[Tite: Comparative Analysis of Vision Transformer (ViT) and ResNet34 on the Roboflow Platform for Differential Classification of Lung Cancer Subtypes Based on CT Scan Images] *Accurate classification of lung cancer subtypes is a crucial step in determining therapy and early diagnosis. However, conventional diagnostic methods based on manual examination still face challenges such as inter-observer variability and time limitations. This study compares the performance of two deep learning architectures, Vision Transformer (ViT) and ResNet34, for classifying four lung cancer subtypes from CT scan images using the Roboflow platform. The dataset consists of 288 CT images divided into four classes: adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cells. All images underwent standard preprocessing and architecture-based augmentation. The evaluation results show that ResNet34 achieved the most optimal performance with an accuracy of 95.7% and an F1-score of 96.3%, as well as better convergence stability. ViT still demonstrated competitive performance with an accuracy of 94.3% and an F1-score of 93.6%, although it was more sensitive to data limitations. Qualitative analysis revealed that misclassifications mainly occurred in ambiguous cases with morphological similarities, particularly between adenocarcinoma and squamous cell carcinoma. In such conditions, ViT more frequently misclassified, while ResNet34, although correct in prediction, produced a low confidence score (0.56), indicating model uncertainty in difficult samples. This study also shows that conventional CNN architectures such as ResNet34 outperform transformer-based architectures under data limitations. The developed models have the potential to be integrated into AI-based clinical decision support systems to accelerate diagnostic analysis.*

Keywords: *Lung Cancer; ViT; ResNet34; Deep Learning; CT Scan*

PENDAHULUAN

Kanker paru-paru merupakan penyebab utama kematian akibat kanker secara global dan membutuhkan proses diagnostik yang cepat, akurat dan konsisten [Wehbe et al., 2024, Chomean et al., 2025]. Metode diagnostik berbasis sitologi dan CT scan masih sangat bergantung pada interpretasi ahli, sehingga rentan terhadap variabilitas antar-pengamat. Perkembangan *artificial intelligence* (AI) menawarkan potensi besar untuk meningkatkan objektivitas dan efisiensi proses diagnostik. Salah satu cabang AI yang paling menjanjikan *adalah deep learning*, yang mampu secara otomatis mengekstraksi pola kompleks dan data medis untuk mendukung keputusan klinis [Chomean et al., 2025].

Kemajuan pesat dalam *deep learning* telah menjadikan *Convolutional Neural Network* (CNN) sebagai arsitektur fundamental yang efektif untuk berbagai tugas klasifikasi dan deteksi sel pada citra medis [Chomean et al., 2025]. Salah satu tonggak penting adalah pengembangan *Deep Residual Learning* (ResNet) oleh [He et al. 2015], yang mengatasi masalah degradasi pelatihan dengan memformulasikan lapisan untuk mempelajari fungsi residual. Pendekatan ini memungkinkan pembangunan jaringan yang sangat dalam (hingga 152 lapisan) yang lebih mudah dioptimalkan, menghasilkan peningkatan akurasi signifikan dalam pengenalan citra [He et al., 2015].

Di sisi lain, kebutuhan akan kecepatan real-time mendorong lahirnya arsitektur *You Only Look Once* (YOLO), yang mereformulasi deteksi objek menjadi masalah regresi tunggal yang terpadu dan sangat cepat [Redmon et al., 2016]. Keberhasilan model turunan YOLO sangat menonjol dalam aplikasi medis: YOLOv11n telah divalidasi untuk klasifikasi 14 jenis sel cairan tubuh dengan tingkat kesepakatan yang hampir sempurna dengan ahli [Chomean et al., 2025], sementara YOLOv8 berhasil diterapkan untuk klasifikasi dan deteksi subtipe kanker paru-paru (seperti *Squamous Cell Carcinoma* (SCC), Adenocarcinoma (ADC), dan *Small Cell Carcinoma* (SCLC)) dalam citra CT [Wehbe et al., 2024].

Selain arsitekstur berbasis konvolusi, inovasi juga datang dari pendekatan non-konvolusional seperti *Vision Transformer* (ViT), yang menunjukkan bahwa ketergantungan pada CNN dapat dihilangkan. ViT mampu menyamai bahkan melampaui kinerja model *state-of-the-art* dalam *benchmark* pengenalan citra ketika dilatih pada set data berskala besar [Dosovitskiy et al., 2021].

Meskipun demikian, sebagian besar penelitian *deep learning* pada citra medis masih menggunakan dataset besar dan seimbang, sehingga belum merepresentasikan kondisi klinis nyata yang umumnya terbatas dan tidak seimbang. Pada dataset kecil, model cenderung mengalami *overfitting* dan penurunan generalisasi, sementara ketidakseimbangan kelas dapat menyebabkan bias terhadap kelas mayoritas dan menurunkan sensitivitas terhadap kelas minoritas. Permasalahan ini semakin krusial pada ViT yang bergantung pada data besar, sehingga performanya berpotensi menurun dibandingkan CNN seperti ResNet34 pada dataset terbatas. Namun, perbandingan sistematis kedua arsitektur dalam kondisi ini masih terbatas, terutama terkait pengaruh augmentasi data dan *confidence threshold*.

Dengan demikian, gap penelitian ini terletak pada kurangnya kajian sistematis mengenai performa arsitektur *deep learning* pada dataset kecil dan tidak seimbang, yang lebih sesuai dengan kondisi klinis nyata. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model klasifikasi berbasis *deep learning* menggunakan arsitektur ResNet34 dan *ViT* pada dataset CT scan paru-paru yang kecil dan tidak seimbang dengan memanfaatkan platform Roboflow. Penelitian ini berkontribusi pada: (1) evaluasi komprehensif performa kedua arsitektur pada dataset terbatas, (2) analisis sensitivitas model terhadap teknik augmentasi data dan *confidence threshold*, serta (3) penyediaan rekomendasi empiris mengenai arsitektur yang paling sesuai untuk implementasi klinis dengan keterbatasan data.

METODE

1. Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen komparatif terhadap dua arsitektur *deep learning* (ViT dan ResNet34) untuk klasifikasi empat subtipe kanker paru berdasarkan citra CT scan.

Rangkaian penelitian meliputi beberapa tahapan utama yaitu pengumpulan data, pra-pemrosesan citra, pelabelan data, augmentasi, pelatihan model, serta evaluasi performa. Seluruh tahapan dirancang untuk memastikan bahwa model mampu belajar dari data dengan efektif meminimalkan bias, serta menghasilkan sistem klasifikasi yang akurat dan dapat digeneralisasi dengan baik terhadap data baru.

2. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle, yaitu dataset *Lung and Colon CT Scan Images* (ID referensi dataset: Kaggle Dataset 2022). Seluruh citra tersedia dalam format JPEG dengan resolusi awal 512 × 512 piksel, dan telah dianonimkan sehingga tidak memuat informasi

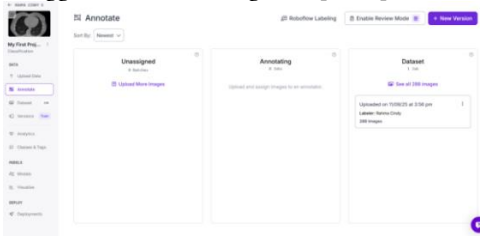
identitas pasien. Dengan demikian, penggunaannya memenuhi prinsip etika dan perlindungan data.

Dataset ini memiliki total 288 citra CT Scan yang terbagi ke dalam empat kelas utama, yaitu adenokarsinoma (120 citra), karsinoma sel besar (51 citra), karsinoma sel skuamosa (90 citra), dan kelas normal (17 citra). Dataset dibagi menjadi *train/validation/test* sebesar 70/20/10 dengan *stratified sampling*. Variasi karakteristik antar kelas terutama mencakup perbedaan morfologi massa tumor, densitas jaringan, dan pola infiltrasi, yang menjadi dasar bagi model untuk mempelajari fitur diagnostik.

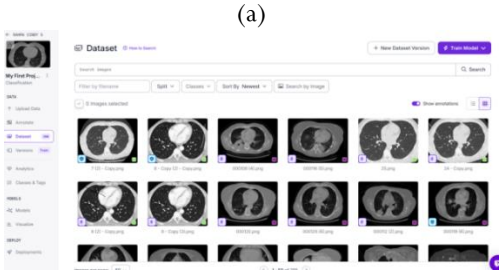
Terdapat ketidakseimbangan jumlah data antar kelas, khususnya karena kelas normal memiliki jumlah yang sangat sedikit dibandingkan kelas lainnya. Untuk memitigasi kondisi *class imbalance* ini, penelitian menerapkan strategi berupa augmentasi tambahan, penataan proporsi data dalam *training set*, serta penerapan *weighted loss p*ada model guna memastikan setiap kelas memperoleh kontribusi pembelajaran yang seimbang. Langkah ini bertujuan meningkatkan kemampuan generalisasi model dan mengurangi bias prediksi terhadap kelas dengan jumlah data dominan.

3. Pelabelan Data

Pelabelan data dilakukan secara manual melalui fitur anotasi pada platform Roboflow. Setiap citra diberi label sesuai kategori sel, yaitu adenokarsinoma, karsinoma sel besar, karsinoma sel skuamosa, dan sel normal. Gambar 3.1 menunjukkan proses pengunggahan dataset ke fitur anotasi. Proses pelabelan mengacu pada deskripsi morfologis yang tercantum dalam dokumentasi dataset Kaggle, sehingga kesesuaian kategori dapat dipertahankan.



(a)



(b)

Gambar 3.1 (a) pengunggahan dataset ke fitur anotasi dan (b) pelabelan citra secara otomatis

4. Pra-Pemrosesan Citra (*Preprocessing*)

Pra-pemrosesan citra dilakukan untuk menyeragamkan format dan kualitas data sebelum pelatihan model. Setiap citra diubah menjadi ukuran 224 × 224 piksel menggunakan metode *stretch* untuk memenuhi kebutuhan arsitektur model sekaligus meningkatkan efisiensi pelatihan. Tahapan pemrosesan meliputi *resizing*, penyesuaian orientasi otomatis, peningkatan kontras menggunakan *contrast stretching* berbasis histogram, serta pemeriksaan kualitas untuk meminimalkan gangguan teknis dan *noise minor*. Proses ini bertujuan menghasilkan citra dengan standar visual yang konsisten sehingga model dapat mempelajari pola morfologi jaringan secara optimal.

5. Augmentasi Data

Augmentasi data digunakan untuk meningkatkan variasi visual secara artifisial sehingga model mampu beradaptasi dengan perbedaan kondisi citra CT Scan di lapangan. Pemilihan teknik augmentasi disesuaikan dengan karakteristik arsitektur masing-masing model karena perbedaan mekanisme ekstraksi fitur antara ViT dan ResNet34.

Untuk klasifikasi menggunakan ViT, augmentasi meliputi rotasi 90°, *flipping horizontal*, dan penyesuaian saturasi −25% hingga +25%. ViT dipilihkan augmentasi ini karena model *transformer* sensitif terhadap distribusi *patch* dan perubahan warna (Dosovitskiy et al., 2021). Oleh karena itu, augmentasi berbasis intensitas penting untuk menjaga konsistensi representasi global tanpa mengubah struktur anatomi, sementara augmentasi spasial diterapkan secara moderat agar tidak mengganggu pola *patch*.

Sementara itu, untuk klasifikasi menggunakan ResNet34, augmentasi yang digunakan mencakup rotasi 90°, perubahan saturasi −25% hingga +25%, penyesuaian paparan −10% hingga +10%, serta *flipping horizontal*. CNN seperti ResNet34 lebih *robust* terhadap variasi orientasi dan pencahayaan karena sifat konvolusi yang mengekstraksi fitur lokal secara stabil. Dengan demikian, augmentasi ini memperluas keragaman tekstur dan pencahayaan tanpa mengurangi kualitas fitur diagnostik.

Perbedaan strategi augmentasi antara kedua model muncul karena ViT lebih sensitif terhadap variasi global, sedangkan ResNet34 lebih stabil terhadap perubahan spasial dan intensitas. Oleh sebab itu, optimasi augmentasi dilakukan sesuai karakteristik masing-masing arsitektur untuk memaksimalkan performa klasifikasi.

6. Pelatihan Model

Pelatihan model dilakukan menggunakan platform dengan 2 arsitektur (*deep learning*) yaitu ViT

Classification dan ResNet34 *Classification*. Dataset hasil pra-pemrosesan dan augmentasi diunggah untuk dilakukan pelatihan dengan parameter-parameter evaluasi yang dipantau pada setiap *epoch*, meliputi: *precision, recall, dan mean average precision (mAP).* Pemantauan metrik dilakukan untuk memastikan model belajar secara stabil, mendeteksi potensi *overfitting*, serta mengoptimalkan kemampuan model dalam mengenali fitur morfologis sel kanker paru-paru.

7. Evaluasi Model

Setelah pelatihan selesai, model dievaluasi menggunakan data uji untuk mengukur performanya dalam mengklasifikasikan citra secara akurat. Evaluasi dilakukan menggunakan beberapa metrik utama, yaitu *akurasi (accuracy), precision, recall, dan F1-score* dihitung berdasarkan jumlah *true positive* (TP, prediksi benar untuk kelas positif*), true negative (*TN, prediksi benar untuk kelas negatif), *false positive* (FP, prediksi salah ketika model mengidentifikasi positif padahal sebenarnya negatif), dan *false negative* (FN, prediksi salah ketika model gagal mendeteksi kelas positif). Keempat metrik tersebut bisa digunakan menggunakan rumus berikut:

- *Accuracy* mengukur proporsi prediksi yang benar terhadap seluruh sampel.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Precision* menilai seberapa tepat model dalam mengidentifikasi kelas positif tanpa menghasilkan terlalu banyak kesalahan positif (false positive).

$$precision = \frac{TP}{TP + FP}$$

- *Recall* mengukur kemampuan model mendeteksi seluruh sampel positif, sehingga penting pada konteks medis yang menekankan minimisasi kesalahan negatif (false negative).

$$Recall = \frac{TP}{TP + FN}$$

- *F1-score* merupakan harmonisasi precision dan recall, dan lebih representatif ketika terjadi ketidakseimbangan jumlah sampel tiap kelas (*class imbalance*).

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall}$$

Evaluasi dilakukan baik secara keseluruhan (*macro, micro, dan weighted average*) maupun per kelas untuk memperoleh gambaran performa yang lebih komprehensif. *Macro average* menghitung rata-rata metrik setiap kelas secara seimbang tanpa memperhatikan jumlah sampel, sehingga penting untuk menilai performa pada kelas minoritas. Sebaliknya, *micro average* menjumlahkan seluruh TP, FP, dan FN antar-kelas sehingga lebih merepresentasikan performa global terutama ketika terjadi ketidakseimbangan jumlah sampel. *Weighted average* turut digunakan untuk menyesuaikan kontribusi tiap kelas berdasarkan proporsi datanya. Visualisasi performa mencakup *confusion matrix*, ROC *curve*, dan *Precision–Recall* (PR) *curve*. *Confusion matrix* membantu memetakan distribusi prediksi dan pola kesalahan antar-kelas, ROC curve menggambarkan kemampuan model membedakan kelas melalui hubungan *true positive rate* dan *false positive rate*, sedangkan PR *curve* menampilkan hubungan precision dan *recall* yang lebih informatif pada kondisi class imbalance. Analisis ini memberikan wawasan mengenai kelas yang paling mudah atau sulit dideteksi oleh model, sehingga dapat dijadikan dasar perbaikan pada pengembangan model selanjutnya.

HASIL DAN PEMBAHASAN

1. Deskripsi Dataset dan Persiapan Data

Dataset yang digunakan dalam penelitian ini berasal dari *platform* Kaggle dan terdiri atas 288 citra CT Scan jaringan paru-paru yang telah dikategorikan ke dalam empat kelas utama, yaitu adenokarsinoma, karsinoma sel besar, karsinoma sel skuamosa, dan sel normal. Adapun distribusi citra pada masing-masing kelas meliputi 120 citra adenokarsinoma, 51 citra karsinoma sel besar, 90 citra karsinoma sel skuamosa, serta 17 citra sel normal. Seluruh citra bersifat anonim dan tidak memuat informasi identitas pasien, sehingga penggunaannya tetap berada dalam koridor etika penelitian dan perlindungan data medis. Dataset ini kemudian diunggah ke *platform* Roboflow untuk melalui tahapan pra-pemrosesan dan persiapan data sebelum digunakan dalam proses pelatihan model.

Tahapan pra-pemrosesan (*preprocessing*) diterapkan untuk menyeragamkan format dan kualitas citra, meliputi perubahan ukuran menjadi 224 × 224 piksel, normalisasi warna dan orientasi, pemeriksaan kualitas citra untuk mengidentifikasi potensi gangguan teknis, serta pengurangan *noise minor* guna memastikan konsistensi visual antar sampel. Setelah itu, dilakukan augmentasi data untuk meningkatkan keragaman representasi visual serta memperkuat kemampuan generalisasi model. Pada arsitektur *Vision Transformer* (ViT) *Classification*, augmentasi mencakup rotasi 90°, *flipping horizontal*, dan modifikasi saturasi sebesar −25% hingga +25%. Sementara itu, pada arsitektur ResNet34 *Classification*, augmentasi yang diterapkan meliputi rotasi 90°, pengaturan saturasi −25% hingga +25%, pengaturan paparan (*exposure*) sebesar −10% hingga

+10%, serta *flipping horizontal*. Penerapan augmentasi ini bertujuan untuk mengatasi ketidakseimbangan jumlah sampel antar kelas dan mensimulasikan variasi kondisi citra CT Scan pada situasi klinis yang lebih luas.

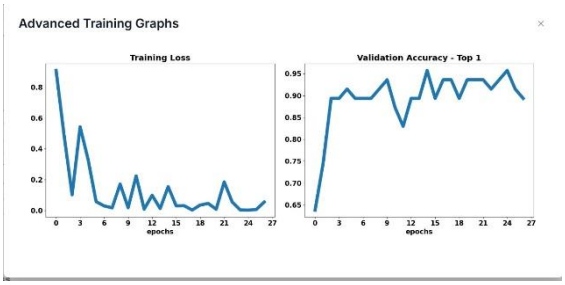
Setelah proses pra-pemrosesan dan augmentasi selesai dilakukan, dataset dibagi ke dalam tiga subset menggunakan rasio 70% (202 citra) sebagai data pelatihan (*training*), 20% (58 citra) sebagai data validasi (*validation*), dan 10% (28 citra) sebagai data pengujian (*testing*). Pembagian ini dimaksudkan untuk menjamin bahwa model memperoleh data latih yang memadai, memiliki mekanisme evaluasi selama pelatihan, serta diuji menggunakan data independen yang tidak pernah diproses sebelumnya.

Dengan demikian, keseluruhan tahapan persiapan data ini memastikan bahwa dataset telah memenuhi standar kualitas dan kuantitas yang diperlukan untuk pelatihan model klasifikasi berbasis *deep learning* pada platform Roboflow 3.0.

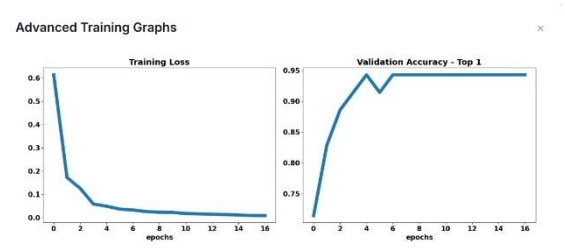
2. Analisis Hasil Pelatihan Model (ViT dan ResNet34)

Pelatihan kedua arsitektur model, yaitu Vision Transformer (ViT) *Classification* dan ResNet34 *Classification*, dievaluasi melalui kurva *training loss* dan *validation accuracy* untuk mengamati dinamika proses pembelajaran serta kestabilan generalisasi model terhadap data validasi. Grafik pelatihan digunakan untuk menilai konvergensi model dari *epoch* awal hingga akhir, serta mendeteksi potensi *overfitting* atau *underfitting* selama pelatihan.

Kurva pelatihan untuk model ResNet34 disajikan pada Gambar 4.1. Grafik tersebut memperlihatkan penurunan *training loss* yang sangat tajam pada *epoch* awal, dari nilai mendekati 0,8 hingga berada di bawah 0,1 hanya dalam empat *epoch* pertama. Pada *epoch-epoch* selanjutnya, nilai loss terus menurun secara bertahap dengan fluktuasi sangat kecil, menunjukkan bahwa proses pembelajaran berlangsung stabil. Sementara itu, kurva akurasi validasi memperlihatkan peningkatan yang cepat pada *epoch* awal, dari sekitar 0,64 hingga melampaui 0,90, kemudian stabil pada rentang 0,90–0,95. Fluktuasi kecil pada akurasi validasi merupakan variasi normal dan tidak menunjukkan adanya penurunan performa model, melainkan mencerminkan sensitivitas arsitektur ResNet34 terhadap variasi sampel validasi pada dataset berukuran terbatas.



Gambar 4.1 Kurva *training loss* dan *validation accuracy* untuk model ResNet34**.**



Gambar 4.2 Kurva *training loss* dan *validation accuracy* untuk model *Vision Transformer* (ViT).

Sementara itu, Gambar 4.2 menampilkan kurva pelatihan untuk model ViT. Grafik menunjukkan penurunan *training loss* yang lebih halus dan stabil dibandingkan ResNet34. Nilai loss menurun secara konsisten dari sekitar 0,6 pada *epoch* pertama hingga mendekati 0,01 pada akhir pelatihan, tanpa adanya peningkatan kembali yang biasanya menjadi indikator ketidakstabilan pembelajaran. Kurva *validation accuracy* pada ViT juga meningkat secara progresif, dari sekitar 0,72 menjadi lebih dari 0,90 pada *epoch* ketiga, dan selanjutnya menetap pada kisaran 0,94 hingga pelatihan selesai. Stabilitas ini mencerminkan kekuatan arsitektur residual dalam mengekstraksi fitur spasial secara efisien pada dataset dengan jumlah sampel terbatas.

Berdasarkan kurva pelatihan pada kedua model, tidak ditemukan indikasi overfitting. Penurunan *training loss* sejalan dengan kestabilan akurasi validasi, dan tidak terdapat divergensi antara keduanya yang biasanya mengindikasikan penurunan generalisasi. Model ResNet34 mencapai kestabilan performa sekitar *epoch* ke-10, sedangkan ViT mencapai kestabilan lebih awal, yaitu sekitar *epoch* ke-5. Hal ini konsisten dengan karakteristik arsitektur, di mana ViT umumnya memiliki proses konvergensi yang lebih cepat pada dataset citra medis yang relatif kecil.

Secara keseluruhan, kedua model menunjukkan kemampuan pembelajaran yang baik. Vision Transformer (ViT) memperlihatkan stabilitas pelatihan yang lebih tinggi, ditandai dengan kurva pelatihan yang lebih konsisten dan fluktuasi yang

lebih kecil selama proses training. Selain itu, ViT juga menunjukkan generalisasi yang lebih baik, yaitu kemampuan model dalam mempertahankan performa pada data validasi tanpa mengalami *overfitting* yang signifikan. Namun, berdasarkan hasil evaluasi akhir pada Tabel 4.1, performa akhir ResNet34 dalam metrik klasifikasi masih lebih tinggi dibandingkan ViT. Oleh karena itu, meskipun ViT unggul dalam stabilitas pelatihan dan generalisasi, ResNet34 menunjukkan keunggulan dalam performa akhir pada dataset yang digunakan.

3. Evaluasi Kinerja Model

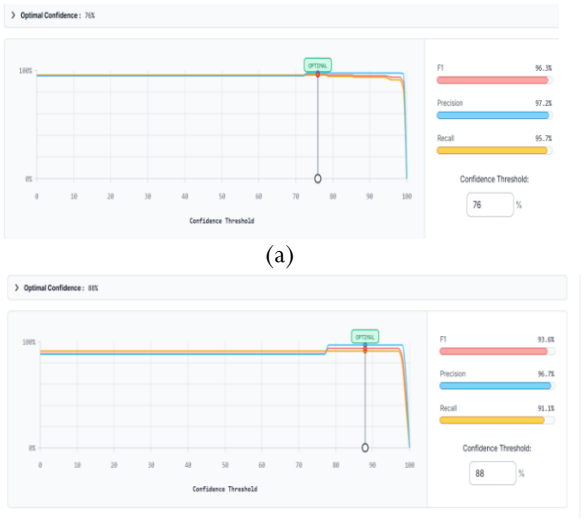
Tahap ini menyajikan evaluasi kinerja akhir dari kedua model klasifikasi, *Vision Transformer* (ViT) dan ResNet34, menggunakan testing set yang independen untuk memastikan kemampuan generalisasi model terhadap citra CT scan kanker paru-paru yang belum pernah dilihat sebelumnya. Berbeda dengan analisis kurva pelatihan yang berfokus pada konvergensi, bagian ini berpusat pada metrik kuantitatif seperti Akurasi, Presisi, Recall, F1_Score, dan analisis *Confusion Matrix* untuk mengidentifikasi kekuatan dan kelemahan masing-masing arsitektur dalam membedakan empat kelas kanker paru-paru. Perbandingan hasil yang diperoleh antara ViT dan ResNet34 akan memberikan wawasan kritis mengenai arsitektur mana yang lebih efektif dalam tugas klasifikasi citra medis yang kompleks ini. Evaluasi Peforma dilakukan melalui fitur *Model Evaluation* pada *Roboflow* yang menggunakan data uji. Metrik yang dianalisis meliputi *accuray, precision, recall, dan F1-score*, dengan *confidence threshold* optimal pada masing-masing model. Hasil evaluasi ditampilkan pada Tabel 4.1.

Tabel 4.1 Perbandingan performa model ResNet34 dan ViT

Model	Precision (%)	Accuray (%)	Recall (%)	F1-Score (%)
ResNet34	97,2	95,7	95,7	96,3
ViT	96,7	94,3	91,1	93,6

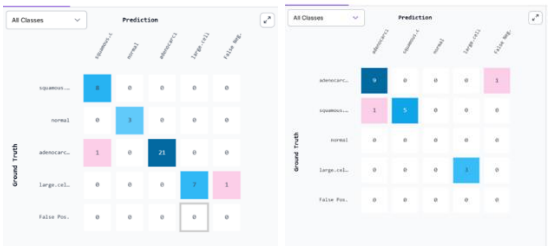
Nilai F1-score yang tinggi pada kedua model menunjukkan keseimbangan antara ketepatan prediksi dan kemampuan model mengenali seluruh kelas. ResNet34 dengan *precision* 97,2% dan *recall* 95,7% mencerminkan tingkat kesalahan prediksi yang sangat rendah. Sementara itu, ViT juga menunjukkan performa kuat dengan *precision* 96,7% dan *recall* 91,1%, menegaskan kemampuannya dalam melakukan generalisasi pada variasi visual.

Kinerja unggul Performa ResNet34 dapat dijelaskan oleh arsitektur residual yang efektif dalam menangkap fitur spasial lokal, menjadikannya lebih adaptif pada dataset terbatas. Sebaliknya, ViT yang mengandalkan *patch-based embedding* dan *self-attention* memerlukan data skala lebih besar untuk mengoptimalkan kemampuan pemahaman konteks global.



Gambar 4.3 (a) Grafik *confidence threshold optimal* pada model VIT *Classification* (76%) dan (b) Grafik *confidence threshold optimal* pada model ResNet34 *Classification* (88%).

Gambar 4.3 merupakan grafik *Confidence Threshold* yang menunjukkan perbedaan signifikan dalam ambang batas keyakinan optimal (*Optimal Confidence Threshold*) antara kedua arsitektur. Model ResNet34 mencapai keseimbangan kinerja terbaik (*F1-Score* 96,3%, *Precision* 97,2%, dan *Recall* 95,7%) pada ambang batas yang lebih rendah, yaitu 76%. Hal ini mengindikasikan bahwa ResNet34 cenderung lebih percaya pada keputusannya meskipun dengan skor probabilitas yang sedikit lebih rendah. Sebaliknya, ViT memerlukan ambang batas keyakinan yang lebih tinggi, yaitu 88%, untuk mencapai kinerja optimal (*F1-Score* 93,6%, *Precision* 96,7%, dan Recall 91,1%). Perbedaan ini menyiratkan bahwa, dalam konteks klasifikasi citra CT scan 4 kelas ini, model ViT memerlukan bukti atau skor probabilitas yang lebih kuat dari mekanisme *self-attention*-nya untuk menganggap suatu prediksi valid, dibandingkan dengan ResNet34 yang lebih liberal.



Gambar 4.4 Analisis *Confusion matrix* pada model ResNet34 *Classification* dan Analisis *Confusion Matrix* pada model ViT *Classification*

Analisis *confusion matrix* (Gambar 4.4) menunjukkan bahwa kesalahan klasifikasi paling sering terjadi antara adenokarsinoma dan sel skuamosa karsinoma, yang memiliki kesamaan morfologi nukleus dan pola distribusi sitoplasma. Namun demikian, tingkat kesalahan tetap berada di bawah 10%, menunjukkan stabilitas sistem klasifikasi yang andal.

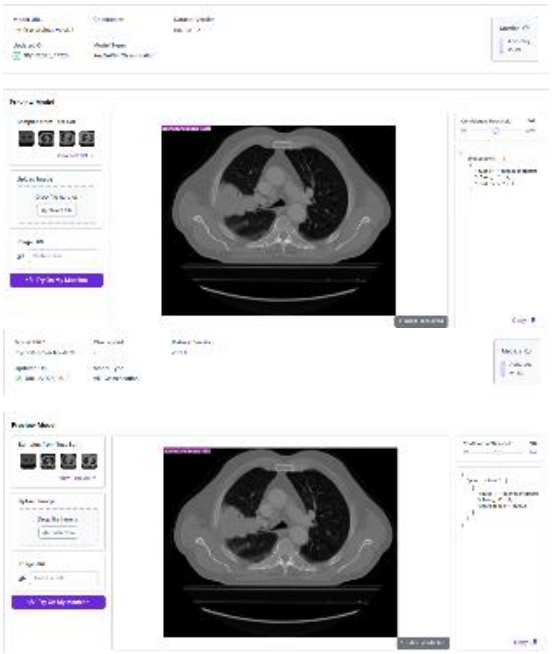
Secara keseluruhan, hasil penelitian menunjukkan bahwa: ResNet34 merupakan model paling optimal dengan akurasi, *precision, recall*, dan F1-score tertinggi. *Vision Transformer* menawarkan kemampuan generalisasi yang kuat dan performa yang tetap kompetitif. Kedua model layak digunakan untuk sistem identifikasi otomatis citra sitologi dan berpotensi dikembangkan dalam aplikasi klinis berbasis AI.

Analisis *confusion matrix* menunjukkan bahwa kesalahan klasifikasi umumnya terjadi antara antara kategori adenokarsinoma dan karsinoma sel skuamosa Kedua kategori ini memiliki kemiripan struktur inti, densitas sitoplasma, dan pola distribusi kromatin, sehingga model membutuhkan representasi fitur yang lebih kaya untuk memisahkannya. Selain itu, kelas sel normal memiliki akurasi terendah akibat jumlah data yang sangat terbatas (17 citra) menyebabkan model kurang mampu mempelajari variasi morfologinya secara komprehensif.

4. Interpretasi Prediksi

Analisis kualitatif kinerja model dilakukan dengan menguji studi kasus spesifik, yaitu melalui penayangan contoh citra CT scan yang berhasil dan gagal diklasifikasikan oleh model. Interpretasi ini bertujuan untuk mengungkap alasan di balik keputusan model, mengidentifikasi apakah kegagalan klasifikasi timbul dari ambiguitas visual antar kelas atau dari keterbatasan arsitektur ViT dan ResNet34.

Berdasarkan analisis *Confusion Matrix*, terlihat bahwa kedua model tersebut hampir akurat dalam melakukan klasifikasi terhadap citra CT Scan dalam 4 kelas. Akurasi tinggi ini didukung dengan pengujian model pada data uji baru yang divisualisasikan untuk studi kasus. Hasil dari prediksi spesifik ini, termasuk contoh *True Positive* dan *False Negative* untuk analisis kesalahan, dapat dilihat pada Gambar 4.5.

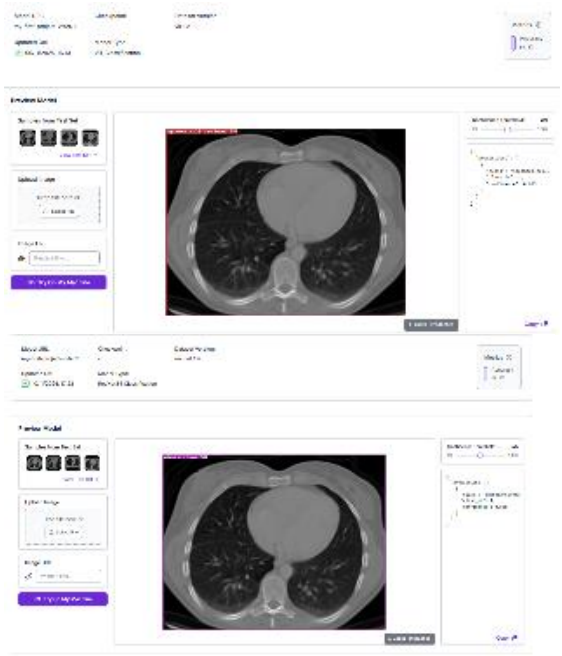


Gambar 4.5. Hasil Prediksi Citra CT Scan Kelas Adenokarsinoma pada model ViT (atas) dan model Resnet34 (bawah).

Studi kasus menunjukkan bahwa ResNet34 lebih konsisten dalam memprediksi kelas sulit, meskipun dengan tingkat keyakinan yang rendah. Pengujian lebih lanjut menggunakan citra uji tambahan memperlihatkan bahwa kedua arsitektur, ViT dan ResNet34, mampu mengenali kelas Adenokarsinoma dengan keyakinan tinggi (0,999 dan 1,000) pada sebagian besar sampel. Hal ini menegaskan bahwa keduanya mampu mengidentifikasi pola morfologis utama yang menjadi ciri khas subtipe tersebut. Namun, analisis rinci pada Gambar 4.6 menunjukkan adanya kasus salah klasifikasi yang signifikan pada arsitektur ViT. Model ViT mengelompokkan citra Adenokarsinoma sebagai Karsinoma Sel Kecil Skuamosa, yang kemungkinan disebabkan oleh kemiripan struktur jaringan antara kedua subtipe, termasuk pola densitas jaringan dan karakteristik inti sel. Kesamaan ini diduga mengganggu mekanisme *self-attention* ViT dalam membedakan fitur-fitur lokal yang bersifat subtil tetapi penting secara diagnostik.

Sebaliknya, ResNet34 berhasil memprediksi kelas sebenarnya, yaitu Adenokarsinoma, meskipun *confidence score* yang dihasilkan relatif rendah (0,56). Nilai keyakinan yang mendekati ambang batas keputusan ini menunjukkan bahwa model mengenali fitur yang benar, tetapi

informasi visual pada citra tersebut tidak cukup dominan atau cenderung ambigu. Dengan demikian, meskipun prediksi ResNet34 lebih akurat, keputusan model tetap disertai tingkat ketidakpastian yang cukup tinggi, mengindikasikan perlunya peningkatan kapasitas model dalam menangani citra dengan kemiripan morfologis tinggi antar subtipe. ViT lebih sering salah identifikasi pada kelas dengan fitur morfologis halus.



Gambar 4.6 Hasil Prediksi Citra CT Scan Kelas Adenokarsinoma pada model ViT (atas) dan model Resnet34 (bawah)

5. Perbandingan dengan Penelitian selanjutnya, Implikasi Klinis dan Keterbatasan

Hasil penelitian ini sejalan dengan studi terdahulu yang menunjukkan efektivitas arsitektur berbasis CNN dan Transformer dalam klasifikasi citra medis. Lin et al. (2021) menunjukkan peforma model Resnet101 dengan akurasi sekitar 95,5% pada klasifikasi sitologi, sementara Dosovitskiy et al. (2021) menunjukkan bahwa model ViT dapat mencapai performa setara atau lebih tinggi dibanding CNN konvensional ketika diberi dataset berukuran besar.

Dengan akurasi di atas 94% pada dataset sitologi berukuran terbatas, studi ini memberikan kontribusi penting bahwa penggunaan Roboflow dan augmentasi yang tepat memungkinkan model deep learning tetap mencapai performa kompetitif meskipun tidak menggunakan dataset berskala besar. Temuan ini sekaligus menguatkan bahwa integrasi validasi ahli merupakan langkah kunci dalam menjaga reliabilitas model.

Penelitian ini memberikan implikasi signifikan dalam konteks diagnostik sitologi. Model ResNet34 dan ViT berpotensi digunakan sebagai sistem pendukung keputusan untuk membantu ahli laboratorium dalam proses identifikasi sel, sehingga dapat mempercepat analisis, meningkatkan konsistensi evaluasi, dan mengurangi beban kerja manual. Namun demikian, penelitian ini memiliki beberapa keterbatasan. Ukuran dataset yang terbatas, khususnya pada kelas sel normal, membatasi kemampuan model dalam melakukan generalisasi pada variasi citra yang lebih luas. Selain itu, seluruh citra berasal dari satu sumber mikroskopis, sehingga variasi alat, metode pewarnaan, dan kondisi pencahayaan belum terwakili secara memadai.

Secara khusus, novelty penelitian ini terletak pada evaluasi komparatif antara arsitektur CNN (ResNet34) dan Vision Transformer (ViT) pada dataset sitologi paru berukuran kecil dan tidak seimbang, yang masih jarang dikaji dalam literatur. Selain itu, penelitian ini juga memberikan kontribusi empiris terkait efektivitas kombinasi augmentasi data dan pengaturan *confidence threshold* dalam meningkatkan performa model pada kondisi data terbatas, serta menunjukkan bahwa pendekatan tersebut tetap dapat menghasilkan performa tinggi yang relevan untuk implementasi klinis.

KESIMPULAN DAN SARAN

Secara keseluruhan, penelitian ini menunjukkan bahwa kedua arsitektur deep learning, ResNet34 dan *Vision Transformer* (ViT), mampu memberikan kinerja yang sangat kompetitif dalam klasifikasi diferensial subtipe kanker paru-paru berbasis citra CT scan. ResNet34 tampil sebagai model dengan performa paling optimal, ditunjukkan oleh akurasi 95,7%, stabilitas konvergensi yang lebih baik, serta efektivitas dalam mengekstraksi fitur spasial lokal. Sementara itu, ViT tetap menunjukkan performa yang kuat dengan akurasi 94,3%, meskipun lebih sensitif terhadap keterbatasan jumlah data.

Kinerja tinggi kedua model, yang didukung oleh nilai F1-score yang seimbang dan penetapan ambang keyakinan (*confidence threshold*) optimal 76% untuk ResNet34 dan 88% untuk ViT mengonfirmasi potensi keduanya untuk diterapkan sebagai sistem pendukung keputusan dalam mempercepat proses analisis klinis. Namun, temuan analisis kualitatif menunjukkan bahwa tantangan tetap muncul pada kasus ambigu, terutama ketika subtipe kanker memiliki kemiripan morfologi. Dalam kondisi tersebut, ViT lebih sering salah mengklasifikasikan, sementara ResNet34 meskipun tepat dalam prediksi, menghasilkan tingkat keyakinan yang rendah (0,56),

mengindikasikan keraguan model terhadap sampel sulit.

DAFTAR PUSTAKA

Ciresan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, 411–418.

Chomean, S., Khemtonglang, N., Mukda, E., & Kaset, C. (2025). AI-powered body fluid cell classification: Development and validation using Roboflow and YOLOv11n framework. *Telematics and Informatics Reports*, *19*, 100243.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, S., Unterthiner, T., … & Houlsby, G. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR)*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Le, T. T., Nguyen-Truong, V.-T., Duong, Q. V. N., Phan, N. T. L., Dao, P. N. T., Mavuso, M. F., … & Quang, K. T. (2025). Deep learning-based classification of colorectal cancer in histopathology images for category detection. Biology Methods and Protocols, 10(1).

Lin, C. K., Chang, J., Huang, C. C., Wen, Y. F., Ho, C. C., & Cheng, Y. C. (2021). Effectiveness of convolutional neural networks in the interpretation of pulmonary cytologic images in endobronchial ultrasound procedures. *Cancer Medicine*, *10*(24), 9047–9057. https://doi.org/10.1002/cam4.4383

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.

Roboflow. (2024). *Roboflow Platform Documentation: Classification Models and Training Overview.*

Sokolova, M., & Lapalme, P. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 420–435.

Sumber Data Kaggle. (n.d.). *Lung Cancer Histopathology Images Dataset.* Repositori data Kaggle. https://www.kaggle.com/datasets/borhanitrash/lung-cancer-ct-scan-dataset

Wehbe, A., Dellepiane, S., & Minetti, I. (2024). Enhanced Lung Cancer Detection and TNM Staging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging. *IEEE Access*, *Volume 6*(2024).