



Implementasi Vision Transformer untuk Klasifikasi Penyakit Pneumonia melalui Citra Chest X-Ray

Raihan Thobie Nabil Maulana¹, Yisti Vita Via^{*2}, Eka Prakarsa Mandyartha³

^{1,2,3}Informatika, Fakultas Ilmu Komputer, UPN Veteran Jawa Timur, Surabaya, Indonesia

Email: ¹raihanthobi@gmail.com; ^{*2}yistivia.if@upnjatim.ac.id; ³eka_prakarsa.fik@upnjatim.ac.id

Maulana, R. T. N., Via, Y. V., & Mandyartha, E. P. (2025). Implementasi Vision Transformer untuk Klasifikasi Penyakit Pneumonia melalui Citra Chest X-Ray. *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 11(2), 249-254

DOI: <https://doi.org/10.33050/cerita.v11i2.3599>

ABSTRAK

Pneumonia salah satu jenis infeksi pernapasan pada saluran pernapasan yang sering disebabkan oleh virus atau bakteri. Kualitas udara yang buruk di wilayah perkotaan Jakarta meningkatkan risiko masyarakat terkena penyakit pneumonia, infeksi saluran pernapasan akut (ISPA), dan asma. pada tahun 2019 menunjukkan bahwa lebih dari 740.180 anak di bawah usia 5 tahun meninggal karena kasus pneumonia, angka tersebut sekitar 14% dari keseluruhan kematian anak usia dini. Dalam mengatasi penyakit pneumonia ini para peneliti medis sudah melakukan banyak penelitian terkait permasalahan diagnosis dini terhadap penyakit pneumonia. Terdapat salah satu tehnik untuk mendeteksi pneumonia yaitu melalui hasil foto Rontgen Dada/*chest x-ray* yang sudah banyak dikembangkan untuk klasifikasi. Vision Transformer (ViT) merupakan salah satu arsitektur Deep Learning yang dikembangkan khusus untuk pengolahan citra. Tujuan penelitian ini untuk mengimplementasikan tugas klasifikasi penyakit pneumonia dengan ViT yang diharapkan dapat membantu mendeteksi dini penyakit pneumonia sehingga dapat penanganan lebih cepat dan lebih baik. Hasil dari penelitian menunjukkan Model ViT memiliki performa yang baik setelah menerapkan beberapa variasi augmentasi, serta stabil baik dalam pelatihan hingga pengujian yang dilakukan. pada Learning Rate kecil yaitu 0.00001 menghasilkan akurasi 80% untuk kasus klasifikasi penyakit pneumonia melalui Citra *Chest X-Ray*

Kata kunci: Pneumonia, Vision Transformer (ViT), Rontgen, Citra

ABSTRACT

Pneumonia is a type of respiratory infection in the respiratory tract that is often caused by viruses or bacteria. Poor air quality in Jakarta's urban areas increases people's risk of developing pneumonia, acute respiratory infections (ARI), and asthma. 2019 showed that more than 740,180 children under the age of 5 died from pneumonia cases, about 14% of all early childhood deaths. In overcoming pneumonia, medical researchers have conducted many studies related to the problem of early diagnosis of pneumonia. One of the techniques to detect pneumonia is through chest x-rays that have been developed for classification. Vision Transformer (ViT) is one of the Deep Learning architectures developed specifically for image processing. The purpose of this study is to implement the classification task of pneumonia with ViT which is expected to help detect pneumonia early so that it can be treated faster and better. The results of the study show that the ViT model has good performance after applying several variations of augmentation, and is stable both in training and testing. at a small Learning Rate of 0.00001, it produces 80% accuracy for the case of pneumonia disease classification through Chest X-Ray Images.

Keywords: Pneumonia, Vision Transformer (ViT), Chest X-Ray, Images

I. PENDAHULUAN

Kualitas udara yang buruk di wilayah perkotaan Jakarta meningkatkan risiko masyarakat terkena penyakit pneumonia, infeksi saluran pernapasan akut (ISPA), dan asma. Fokus utama pemerintah saat ini adalah pada pneumonia, terutama karena mayoritas kasus terjadi pada balita. Di DKI Jakarta kasus ISPA meningkat sekitar 90% akibat polusi yang kian memburuk setiap harinya hingga pernah menjadi kota nomor 1 dengan kualitas udara terburuk di Dunia (WHO, 2022). Mengingat urgensi masalah ini, pemerintah bertujuan untuk mengurangi angka kasus ISPA yang dapat menyebabkan risiko pneumonia dengan mengedukasi masyarakat tentang pentingnya mengantisipasi dan melindungi diri dari risiko penyakit ini, khususnya melalui program vaksinasi sebagai perlindungan lebih untuk masyarakat.

Pneumonia merupakan salah satu jenis infeksi akut pada saluran pernapasan yang sering disebabkan oleh virus atau bakteri (Andika et al., 2019). Penyakit ini dapat mengakibatkan gejala ringan hingga kondisi yang mengancam jiwa, terutama pada berbagai kelompok usia. Berdasarkan data WHO pada tahun 2019 menunjukkan bahwa lebih dari 740.180 anak di bawah usia 5 tahun meninggal karena pneumonia, angka tersebut sekitar 14% dari keseluruhan kematian anak usia dini. Pneumonia merupakan salah satu dari sepuluh penyakit dengan jumlah kasus terbanyak. Kondisi ini memiliki tingkat kekritisitas yang tinggi karena berpotensi menyebabkan peradangan pada

alveoli di satu atau kedua paru-paru. Serangan pneumonia umumnya terjadi pada anak-anak, populasi lanjut usia, serta individu dengan riwayat masalah kesehatan sebelumnya. Pada kondisi pneumonia, alveoli yang merupakan kantong kecil dalam paru-paru yang berperan dalam pertukaran gas, terisi dengan nanah dan cairan, mengganggu proses pernapasan dan penyerapan oksigen. Penularan infeksi ini biasanya terjadi melalui kontak langsung dengan individu yang terinfeksi. Infeksi pneumonia sering disebabkan oleh bakteri yang merespon infeksi virus tertentu seperti demam atau flu. Beberapa faktor lain juga dapat memperburuk kondisi penyakit ini secara kronis, termasuk riwayat perokok aktif dan pasif, paparan polusi udara, dan faktor-faktor lingkungan lainnya (WHO, 2022).

Dalam mengatasi penyakit pneumonia ini tentunya para peneliti medis sudah melakukan banyak penelitian terkait permasalahan diagnosis dini terhadap penyakit pneumonia (Andromeda Anwar, 2022). Terdapat salah satu tehnik untuk mendeteksi pneumonia yaitu melalui hasil foto Rontgen Dada/chest x-ray yang sudah banyak dikembangkan untuk klasifikasi, proses ini dilakukan dengan mengambil gambar dari Rontgen Dada yang kemudian dilakukan proses identifikasi melihat adanya peradangan (pneumonia) atau tidak dengan memanfaatkan arsitektur Machine Learning untuk memprediksi informasi dari gambar X-Ray (Zaid Munantri et al., 2019). Berdasarkan penjelasan tersebut bahwa untuk mendeteksi pneumonia menggunakan media citra X-Ray itu dapat dilakukan, oleh karena itu dibutuhkan

pengembangan teknologi Machine Learning yang dapat membantu mendiagnosis pneumonia berdasarkan gambar x-ray paru. (Ukwuoma et al., 2023).

Deep Learning merupakan cabang dari Machine Learning yang menggunakan jaringan saraf tiruan (Deep Neural Networks) untuk memahami dan memodelkan data yang kompleks (Roihan et al., 2019). Vision Transformer (ViT) merupakan salah satu arsitektur Deep Learning yang dikembangkan khusus untuk pengolahan citra. ViT menggunakan pendekatan transformer yang awalnya diperkenalkan untuk pengolahan bahasa alami (Dosovitskiy et al., 2021). Pendekatan transformer memungkinkan ViT untuk mempelajari representasi citra secara hierarki dengan membagi citra menjadi potongan-potongan kecil yang disebut "patch" dan menyajikan informasi dari patch tersebut dengan transformasi self-attention. Dengan demikian, ViT dapat mengambil keuntungan dari kapasitas representasi yang lebih kuat dan kemampuan untuk menangkap konteks global dari citra (Singh et al., 2024). Oleh sebab itu hasil yang menjanjikan muncul dari arsitektur ViT ketika menerapkannya kedalam pengolahan citra.

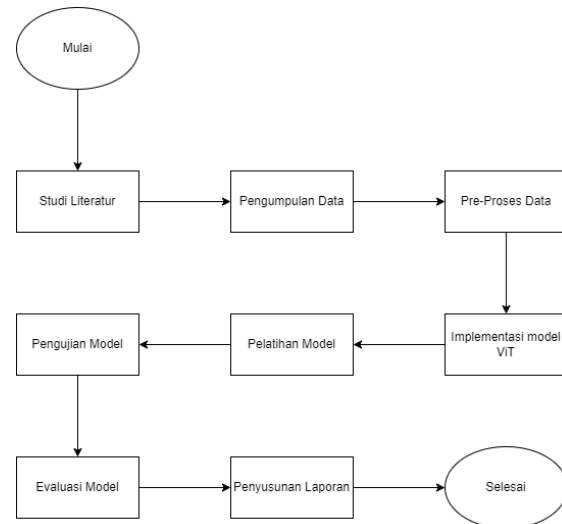
Vision Transformer merupakan sebuah arsitektur yang sepenuhnya didasarkan pada vanilla Transformer. Arsitektur Transformer telah menjadi sangat populer dalam beberapa tahun terakhir, terutama dalam bidang terjemahan mesin dan tugas-tugas pemrosesan bahasa alami (NLP), karena telah menunjukkan kinerja yang sangat baik (Chen et al., 2024). Transformer mengikuti arsitektur encoder-decoder dan mampu memproses data berurutan secara paralel tanpa bergantung pada jaringan berulang. Keberhasilan model Transformer sebagian besar disumbangkan oleh mekanisme Self-Attention, yang dirancang untuk menangkap hubungan spasial antara elemen-elemen dalam suatu urutan (Pangestu et al., 2024).

Berdasarkan latar belakang yang sudah disampaikan, serta penelitian sebelumnya, disini penulis ingin melakukan penelitian yang berjudul "Penerapan Metode Vision Transformer Untuk Proses Klasifikasi Penyakit Pneumonia Melalui Citra Chest X-Ray", untuk mengetahui performa hasil klasifikasi dari model transformer

yaitu ViT pada kasus paru-paru normal dan paru-paru pneumonia.

II. METODE PENELITIAN

Tahapan penelitian yang dilakukan penulis dapat dilihat pada alur dibawah.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Data yang digunakan pada penelitian diambil dari *Open Source Kaggle* yang bernama "*Chest X-Ray Images (Pneumonia)*" dengan jumlah dataset 5856 data citra dari dua kelas, Pneumonia dan Normal. Pembagian data bisa dilihat pada Tabel 1.

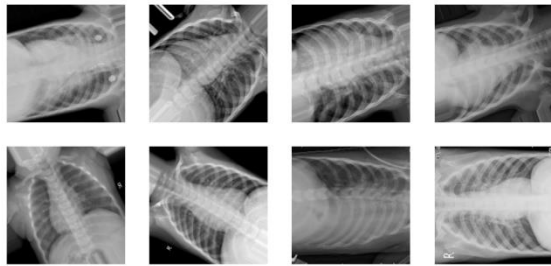
Tabel 1. Dataset *Chest X-Ray*

Kelas	Kategori		Jumlah
	Training	Testing	
Normal	1349	234	1583
Pneumonia	3883	390	4273

Pembagian data atau *splitting* sudah dilakukan sebelumnya pada data yang tersedia dari Kaggle.

B. Pre-Processing Data

Dataset akan dilakukan *pre-processing* yaitu augmentasi untuk memperbanyak variasi citra sehingga model dapat menangkap pola-pola dari keberagaman bentuk citra dan memaksimalkan pelatihan yang dilakukan.



Gambar 2. Hasil Augmentasi Citra

Gambar 2. Menampilkan hasil dari augmentasi yang dilakukan seperti rotasi, *flip*, dan *zoom*. Serta *resizing* yang mana bertujuan untuk merubah ukuran citra menjadi satu ukuran tetap agar tidak terdapat keberagaman ukuran citra input pada saat diproses oleh model.

C. Perancangan Model Vision Transformer (ViT)

Dalam Model ViT akan memproses citra input yang telah diaugmentasi menjadi potongan-potongan kecil yang disebut *patch*, tiap *patch* yang dihasilkan akan diratakan atau *flatten* menjadi ukuran 1 dimensi dan diberikan informasi posisi dari tiap *patch* (satu dimensi) tersebut agar model dapat memahami hubungan antar *patch* melalui mekanisme *Self-Attention* berdasarkan informasi posisi yang sudah diberikan pada tiap *patch* yang ada. Berdasarkan hasil dari *Self-Attention* maka model dapat memahami perbedaan pola atau karakter dari citra input yang telah diolah sehingga dapat melakukan tugas klasifikasi.

D. Pelatihan Model ViT

Pelatihan model akan diuji dengan beberapa parameter *learning rate* dan *Epoch* 50. Tujuan pengujian ini untuk mengetahui performa terbaik dari Model ViT yang bisa dihasilkan berdasarkan *learning rate* dengan keterbatasan data yang ada pada proses pelatihan. Hasil pelatihan terbaik akan dilihat berdasarkan akurasi model yang didapatkan.

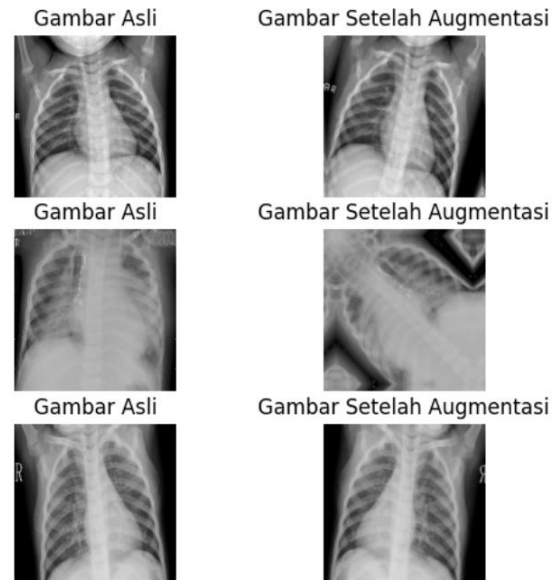
E. Pengujian dan Evaluasi Model ViT

Pengujian model dilakukan untuk mengetahui seberapa baik performa yang dihasilkan oleh Model ViT untuk proses klasifikasi pneumonia melalui citra *Chest X-Ray* dengan menampilkan hasil prediksi dari model serta Evaluasi terkait nilai akurasi, presisi dengan *Confusion Matriks*.

III. HASIL DAN PEMBAHASAN

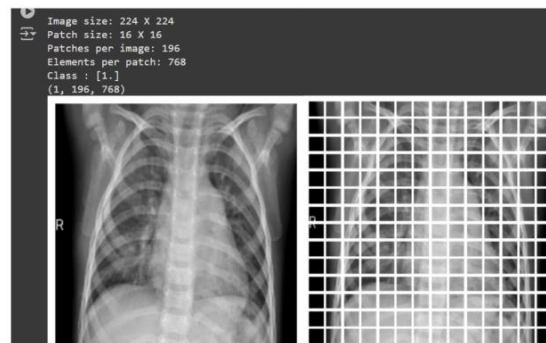
A. Visualisasi Citra

Citra yang digunakan yaitu Paru Pneumonia dan Paru Normal, berikut hasil dari *pre-processing*.



Gambar 3. Hasil *Pre-Processing* Citra

Kemudian citra yang dihasilkan dari lapisan augmentasi akan diproses oleh model dengan membagi citra input yang berukuran 224x224 menjadi *patch* berukuran 16x16 seperti pada Gambar 4.



Gambar 4. Proses *Patch Extraction*

Setelah itu citra akan diproses oleh model dengan memperhatikan pola dari tiap *patch* yang dihasilkan untuk menangkap karakter dan detail dari perbedaan antara kedua kelas yang akan dilakukan klasifikasi.

B. Hasil Implementasi Model ViT

Hasil yang didapatkan dari implementasi Model ViT dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengujian *Learning Rate*

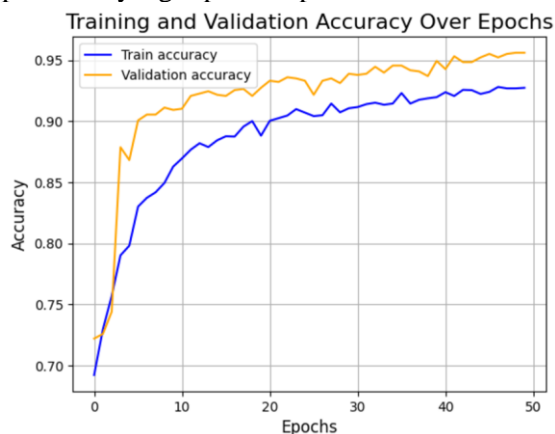
Learning Rate	Akurasi
10^{-2}	62%
10^{-3}	77%
10^{-4}	79%
10^{-5}	80%

Berdasarkan hasil pengujian yang telah dilakukan didapatkan hasil seperti pada Tabel 2. Yang menunjukkan akurasi terbaik pada *learning rate* terendah yaitu 10^{-5} dengan kestabilan terbaik diperoleh pada ukuran *learning rate* tersebut, berikut hasil evaluasi yang didapatkan.

	precision	recall	f1-score	support
Normal	0.91	0.53	0.67	234
Pneumonia	0.78	0.97	0.86	390
accuracy			0.80	624
macro avg	0.84	0.75	0.77	624
weighted avg	0.82	0.80	0.79	624

Gambar 5. Hasil Evaluasi Model ViT

Didapatkan performa terbaik pada akurasi sebesar 80 di Model ViT serta berikut history pelatihan yang diperoleh pada Gambar 6.



Gambar 6. History Pelatihan Model ViT

Berdasarkan hasil history yang didapatkan, model berhasil memperoleh akurasi pelatihan yang tinggi dengan kestabilan selama proses pelatihannya hingga mendapatkan akurasi test pada 80%.

Dengan hasil prediksi yang diperoleh dapat dilihat pada Gambar 7.

```

39/39 ————— 2s 21ms/step
hasil prediksi:
[[0.99886996]
 [0.9917201 ]
 [0.9999964 ]
 [0.05052334]
 [0.97077775]]
hasil prediksi (biner):
[[1]
 [1]
 [1]
 [0]
 [1]]
label asli:
[[1.]
 [1.]
 [1.]
 [1.]
 [1.]]
    
```

Gambar 7. Hasil Prediksi Klasifikasi

Didapatkan hasil prediksi dari 5 data pengujian, 4 data berhasil diprediksi benar dengan 1 prediksi salah.

IV. KESIMPULAN

Metode ViT untuk tugas Klasifikasi Penyakit Pneumonia pada Paru melalui Citra Chest X-Ray mendapatkan hasil yang baik. Berdasarkan Prediksi yang diperoleh pada Pengujian Model dari 5 data yang diprediksi menghasilkan 4 Prediksi Benar untuk klasifikasi yang dilakukan antara Paru Pneumonia dan Paru Normal.

Penerapan nilai Learning Rate yang lebih kecil membuat model ViT lebih stabil selama proses pelatihan dan meningkatkan Test Akurasi yang dihasilkan karena generalisasi model dapat belajar dengan baik.

DAFTAR PUSTAKA

- [1]. Andika, L. A., Pratiwi, H., & Handajani, S. S. (2019). KLASIFIKASI PENYAKIT PNEUMONIA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DENGAN OPTIMASI ADAPTIVE MOMENTUM *. In Indonesian Journal of Statistics and Its Applications (Vol. 3, Issue 3).
- [2]. Andromeda Anwar, R. (2022). Gambaran Tentang Pengaruh Musim Terhadap Kejadian Pneumonia Anak di Kabupaten Natuna Tahun 2020-2021. SURABAYA BIOMEDICAL JOURNAL, 2(1).

- [3]. Chen, T., Philippi, I., Phan, Q. B., Nguyen, L., Bui, N. T., daCunha, C., & Nguyen, T. T. (2024). A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images. *Healthcare Analytics*, 5. <https://doi.org/10.1016/j.health.2024.100332>
- [4]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. <http://arxiv.org/abs/2010.11929>
- [5]. Pangestu, A., Purnama, B., & Risnandar, R. (2024). Vision Transformer untuk Klasifikasi Kematangan Pisang. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 75–84. <https://doi.org/10.25126/jtiik.20241117389>
- [6]. Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
- [7]. Singh, S., Kumar, M., Kumar, A., Verma, B. K., Abhishek, K., & Selvarajan, S. (2024). Efficient pneumonia detection using Vision Transformers on chest X-rays. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-52703-2>
- [8]. Ukwuoma, C. C., Qin, Z., Belal Bin Heyat, M., Akhtar, F., Bamisile, O., Muaad, A. Y., Addo, D., & Al-antari, M. A. (2023). A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *Journal of Advanced Research*, 48, 191–211. <https://doi.org/10.1016/j.jare.2022.08.021>
- [9]. WHO. (2022, November 11). Pneumonia in children. World Health Organization.
- [10]. Zaid Munantri, N., Sofyan, H., & Yanu, M. (2019). Aplikasi Pengolahan Citra Digital Untuk Identifikasi Umur Pohon. In *Telematika (Vol. 16, Issue 2)*.