

Perbandingan Kemampuan Klasifikasi Citra *X-ray* Paru-paru menggunakan *Transfer Learning* ResNet-50 dan VGG-16

Performance Comparison of X-ray Image of Lung Classification using ResNet-50 and VGG-16 Transfer Learning

Tasya Berliani^{1*}, Enggalwiguno Rahardja¹, Lina Septiana¹

¹Universitas Kristen Krida Wacana

Jalan Tanjung Duren Raya no. 4, Jakarta Barat 11470

*Penulis korespondensi

Email: tasya.212019007@civitas.ukrida.ac.id

Received: January 23, 2023

Accepted: May 8, 2023

Abstrak

Di masa pandemi Covid-19, foto rontgen menjadi umum digunakan untuk memeriksa pasien diduga Covid-19. Pada citra *x-ray* paru-paru yang terkena Covid-19 ditemukan adanya bercak putih atau flek; Namun, paru-paru yang memiliki flek ini tidak selalu disebabkan oleh Covid-19. Tujuan penelitian ini adalah untuk mengklasifikasikan beberapa jenis penyakit paru-paru dari citra *x-ray*, yaitu paru-paru dengan Covid-19, paru-paru dengan pneumonia, dan paru-paru yang memiliki flek dibandingkan dengan yang normal. Proses klasifikasi data pada penelitian ini dilakukan dengan membandingkan dua model yaitu CNN VGG-16 dan ResNet-50 dengan model yang telah dilatih sebelumnya. Metrik evaluasi yang digunakan dalam penelitian ini terdiri dari akurasi, presisi, sensitivitas, spesifisitas, skor F1, dan kecepatan waktu inferensi. Hasil menunjukkan bahwa VGG-16 lebih unggul dari ResNet-50 dalam hal kecepatan inferensi namun tidak dalam hal metrik evaluasi lainnya. Perubahan parameter juga menunjukkan hasil yang berbeda; epoch 200 adalah nilai optimal. Untuk mendapatkan hasil yang optimal diperlukan *finetuning* dengan menyesuaikan kondisi data yang digunakan. Sebagai simpulan, VGG-16 memiliki kemampuan klasifikasi yang lebih baik dibandingkan ResNet-50, namun perlu terus dikembangkan dengan memperbanyak data klinis yang aktual.

Kata kunci: Covid-19; klasifikasi citra medis; *x-ray* dada, ResNet-50; VGG-16

How to Cite:

Berliani T, Rahardja E, Septiana L. Perbandingan kemampuan klasifikasi citra *x-ray* paru-paru menggunakan transfer learning resnet-50 dan vgg-16. Journal of Medicine and Health. 2023; 5(2): 123-35. DOI: <https://doi.org/10.28932/jmh.v5i2.6116>

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 

Research Article

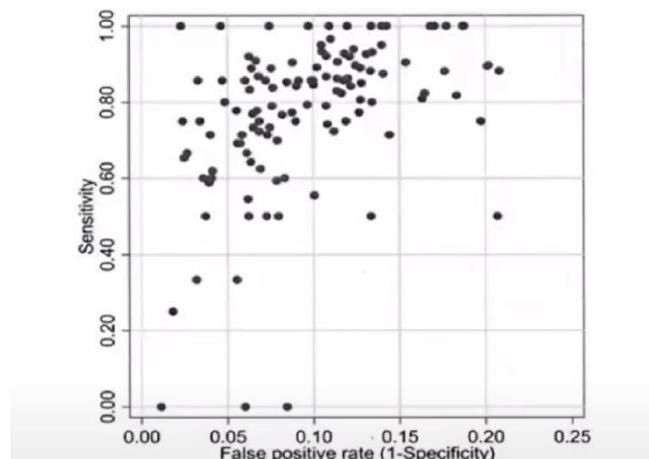
Abstract

During the Covid-19 pandemic, x-ray images became commonly used to examine Covid-19 suspect. X-ray images of the lungs affected by Covid-19 become cloudy; however, the cloudy lung is not always caused by Covid-19. Therefore, this experiment is conducted to classify some kinds of lung diseases from x-ray images, i.e., lung with Covid-19, lung with pneumonia, and lung opacity compared to the normal ones. The data classification process in this experiment was carried out by comparing two models, i.e., CNN VGG-16 and ResNet-50, with pre-trained model. The evaluation matrices used in this experiment consist of accuracy, precision, sensitivity, specificity, F1-score, and speed of inference time. The results show that the VGG-16 is superior to ResNet-50 in speed but not for other evaluation matrices. Parameter changes also show different results; epochs 200 is the optimal value. Finetuning using the actual data is necessary to get the optimum results. In conclusion, VGG-16 has better classification capabilities than ResNet-50, but it needs to be continuously developed by increasing actual clinical data.

Keywords: Covid-19; medical image classification; ResNet-50; VGG-16

Pendahuluan

Citra medis merupakan intervensi medis untuk mempresentasikan secara visual fungsi-fungsi dari organ maupun jaringan yang ada dalam tubuh seseorang. Seiring dengan perkembangan zaman, alat-alat medis juga semakin berkembang seperti terciptanya tabung sinar-X, *Magnetic Resonance Imaging* (MRI), *Computed Tomography* (CT), dan masih banyak lagi. Hal ini dikarenakan alat-alat ini sering digunakan oleh para tenaga medis, maka pengolahan secara otomatis dari citra medis ini semakin diperlukan. Saat ini *gold standard* diagnosis citra radiologi adalah dengan pembacaan secara langsung oleh *expert radiologist*. Efektivitas terkait diagnosis oleh radiologis sudah dibahas oleh Kim.¹ Diagnosis secara manual ini sering kali tidak tepat dan performa dari para radiologis bervariasi karena terdapat subjektivitas dan perbedaan/bias *intra-observer* dan *inter-observer*. Gambar 1 menunjukkan bahwa dari hasil survei performa radiologis terdapat beberapa kasus yang mana hasil pembacaan dari radiologis merupakan hasil *false positive*.²



Gambar 1 Hasil Survei Performa Radiologis.²

Research Article

Alat-alat medis ini menghasilkan gambar yang cukup banyak terutama untuk CT dan juga MRI. Jumlah gambar yang banyak ini memerlukan inspeksi yang lebih teliti.³ Selain itu volume pasien lebih banyak, variasi pemeriksaan lebih banyak sehingga ada kesenjangan jumlah dokter dan hasil pemeriksaan. Untuk membantu proses diagnosis, maka dikembangkanlah beberapa model *Convolutional Neural Network* (CNN) yakni salah satu metode kecerdasan buatan (*Artificial Intelligence*) yang dapat membantu para tenaga kesehatan untuk dapat mengevaluasi gambar medis dengan optimal. Model-model CNN yang dapat digunakan untuk melakukan klasifikasi maupun segmentasi adalah LeNet-5.⁴ Selain LeNet-5, ada AlexNet yang mirip dengan LeNet-5 namun lebih akurat dan *error* nya lebih kecil jauh dibandingkan LeNet-5. Perangkat ZFNet⁵ dan arsitektur dari ZFNet menyerupai AlexNet⁶ namun memiliki perbedaan dalam *hyperparameter*. Model CNN lainnya adalah GoogleNet yang dibuat oleh Google. Selain model yang disebutkan di atas, terdapat model yang juga sering digunakan yaitu VGGNet dan ResNet.

VGGNet dan ResNet dipilih dalam penelitian ini, sebab seperti dalam penelitian sebelumnya telah dijelaskan bahwa kedua arsitektur ini cukup efektif untuk klasifikasi citra *x-ray*⁷ selain dapat melakukan klasifikasi, segmentasi maupun manipulasi pada sebuah gambar. Berbagai jenis citra medis maupun penyakitnya dapat diproses dan dianalisis menggunakan model CNN ini dengan melakukan pengaturan pada parameter-parameter yang ada. Tulisan ini akan membahas mengenai kemampuan ResNet-50 dan VGG-16 dalam melakukan klasifikasi citra medis dalam bentuk *x-ray* yang diambil pada bagian dada secara khusus menjadi empat kategori yaitu kondisi normal, COVID 19, pneumonia, dan lung *opacity*. Adapun penelitian serupa yang sudah ada sebelumnya hanya membahas perbedaan kemampuan ResNet-50 dan VGG-19 dalam melakukan klasifikasi *x-ray* dengan obyek klasifikasi yang berbeda yaitu normal dan pneumonia saja.⁷

Dari *data set* yang sudah dikumpulkan, dilakukan pelatihan/*training* pada sistem untuk dapat mengklasifikasikan data ke dalam 4 kelas yaitu citra *x-ray* normal, citra *x-ray* pada penderita Covid-19, citra *x-ray* untuk *viral* pneumonia, dan citra *x-ray* untuk lung *opacity*. Kelas-kelas ini diambil berdasarkan fakta bahwa hasil citra *x-ray* seringkali menyebabkan kekeliruan diagnosis penyakit Covid-19 ini. Dalam masa pandemi Covid-19 ini, berdasarkan hasil citra *x-ray* yang menunjukkan bercak putih pada paru-paru akan disimpulkan bahwa orang yang memiliki paru-paru yang terdapat bercak putih tersebut menderita Covid-19. Penderita Covid-19 banyak yang mengalami pneumonia dan pneumonia tersebut menimbulkan bercak putih pada paru-paru. Saat orang yang bersangkutan tersebut diperiksa secara mikrobiologi, hasil menunjukkan bahwa orang tersebut negatif Covid-19. Bercak putih pada paru-paru sulit dibedakan apakah akibat *viral* pneumonia atau Covid-19 karena kedua penyakit memberikan citra *x-ray* paru-paru yang serupa^{8,9}

Research Article

Dikarenakan hasil citra medis memiliki tingkat kemiripan yang tinggi, diagnosis secara pasti apakah orang tersebut menderita Covid-19 atau tidak menjadi tantangan cukup besar dalam dunia medis saat ini. Dalam tulisan ini akan dibahas klasifikasi dalam beberapa kelas untuk citra medis yaitu normal, covid, *viral pneumonia* dan *lung opacity* dengan menggunakan ResNet-50 dan VGG-16. Menurut Ismael dan Şengür 2021, ResNet-50 menunjukkan hasil akurasi yang paling tinggi untuk melakukan klasifikasi dibandingkan model lain (VGG16, VGG19, ResNet18, dan ResNet101) yaitu sebesar 92,63%.¹⁰ Alat VGG-16 merupakan model yang hanya memiliki 16 layer yang dimana model nya tidak terlalu berat dan intervensi waktu nya lebih sedikit. Tujuan penulisan ini adalah membandingkan hasil klasifikasi dari VGG-16 dan ResNet-50 dan mengetahui adakah perbedaan yang signifikan antara hasil ResNet-50 dan VGG-16.

Metode

Penelitian ini dilakukan dengan mengambil dua *data set* publik dengan lisensi CC BY-SA 4.0 yang bersumber dari University of Montreal, dan Qatar University berkolaborasi dengan University of Dhaka Bangladesh.^{11,12} Penelitian ini menggunakan *data set* publik dan tidak menggunakan *data set* privat, supaya bisa digunakan sebagai pembanding yang adil apabila di masa depan ada peneliti yang ingin mereproduksi atau mengembangkan penelitian ini. Data yang digunakan tersebut sudah dilengkapi label oleh *expert* radiologis ke dalam empat kelas yaitu normal, covid, *viral pneumonia*, serta *lung opacity*; dan didapatkan dua *data set* yang berbeda. Pada *data set* yang pertama terdapat ke empat kelas tersebut. Namun untuk *data set* yang kedua hanya memiliki citra medis untuk *viral pneumonia*, covid, dan normal serta memiliki jumlah data yang lebih sedikit dibandingkan *data set* pertama. Oleh karena itu *data set* kedua digunakan sebagai *data set* untuk validasi dengan memindahkan 90 data *lung opacity* dari *data set* pertama. Pemilihan jumlah *data set* untuk setiap kelas tidak diacak secara keseluruhan, namun jumlahnya dalam setiap kategori disesuaikan agar proses *training* proporsional dan tidak terjadi *overfitting* pada satu kategori atau kelas tertentu saja. Pada penelitian ini *data set* dibagi 90% untuk *train* dan 10% data untuk *test*.

Pelaksanaan penelitian dimulai dengan menyiapkan *data set*, dalam hal ini gambar yang digunakan diubah ukurannya menjadi 299 x 299 piksel. Data tersebut kemudian diaugmentasi untuk memperbanyak variasi. Selanjutnya, dilakukan proses *training* menggunakan dua metode yaitu VGG-16 dan Resnet-50. *Training* menghasilkan dua buah model yang kemudian digunakan untuk proses pengetesan atau untuk memprediksi data yang baru. Diagram proses tersebut secara garis besar ditunjukkan pada gambar 2.

Research Article



Gambar 2 Diagram Pelaksanaan Penelitian

Penelitian ini dilakukan dengan menggunakan *googlecolab* dengan *runtime* jenis GPU standar dan menggunakan *framework* Keras versi 2.8.2 dan TensorFlow versi 2.8.0. Kami menggunakan beberapa *python library* seperti tensorflow, keras, numpy, pandas, matplotlib.pyplot, seaborn, PIL, cv2, os, tqdm, zipfile, os, datetime dan sklearn. Pada penelitian kali ini, kami menggunakan *Image Data Generator* yang bertujuan untuk memudahkan mengaugmentasi gambar dan kami juga menggunakan *adam optimizer* dengan *learning rate* yaitu 0,002. *Adam optimizer* ini merupakan algoritma yang digunakan untuk mengoptimalkan *error* agar semakin kecil.

1.1. Deskripsi *Data set*

Data yang digunakan dalam penelitian ini dibagi menjadi empat kelas gambar *x-ray* dengan kondisi normal, Covid-19, *viral pneumonia*, dan *lung opacity*. Adapun untuk proses *training*, validasi, *testing* dalam penelitian ini masing-masing menggunakan data yang berbeda. Hal ini dilakukan untuk memastikan bahwa model yang dihasilkan dapat digunakan pada data yang baru. Tabel 1 menunjukkan rincian distribusi data yang digunakan dalam penelitian ini.

Data set Covid-19_Radiography_Data set memiliki data untuk Covid-19 sebanyak 3.616 gambar, *lung opacity* sebanyak 6.012 gambar, *viral pneumonia* sebanyak 1345 dan normal sebanyak 10.192 gambar dengan resolusi 299 x 299. Adapun ukuran resolusi ini adalah ukuran yang umum digunakan dalam proses pelatihan model CNN⁴⁻⁷ yakni sekitar 64 x 64 sampai dengan 400 x 400. Ukuran resolusi ini diambil dengan mempertimbangkan supaya akurasi tetap baik dan kecepatan pemrosesan tidak terlalu lama. Untuk *data set Covid-19 Image Data set* memiliki 137 gambar untuk Covid-19, 70 gambar untuk normal dan 90 untuk *viral pneumonia*. Dari *data set Covid-19 Radiography* dikurangi 90 data *lung opacity*, sisa data dialokasikan sebanyak 80% untuk *training* dan 20% untuk *testing*.

Tabel 1 Distribusi Data

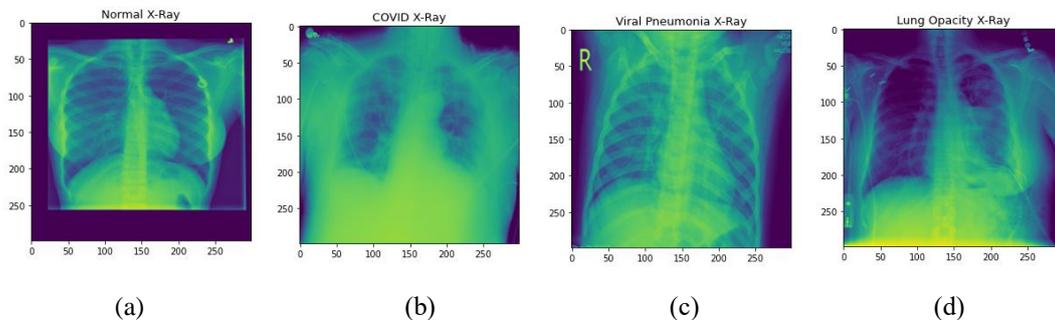
<i>Disease</i>	<i>Train</i>	<i>Test</i>	<i>Validation</i>
Normal	8154*	2083*	90**
Covid-19	2893*	723*	137**
<i>Viral Pneumonia</i>	1076*	269*	90**
<i>Lung Opacity</i>	4738*	1184*	90*

* Sumber dari Covid-19_Radiography_Data set

** Sumber dari Covid-19 Image Data set

Research Article

Gambar 3 menunjukkan contoh citra *x-ray* dari masing-masing kelas, secara berurutan (a) normal, (b) Covid-19, (c) *viral pneumonia*, dan (d) *lung opacity*.



Gambar 3 Contoh Gambar dari Masing-masing Kelas

1.2. Persiapan Data

Kualitas gambar dari *data set* yang diperoleh sudah bagus maka tidak diperlukan pengaturan kualitas gambar. *Rescaling* dan normalisasi data tetap dilakukan untuk memperbanyak variasi data agar mendapatkan model yang lebih baik. Tabel 2 menunjukkan parameter yang digunakan dalam proses augmentasi data.

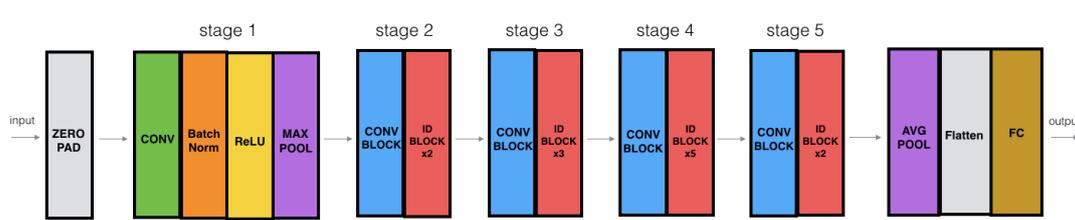
Parameter yang digunakan untuk *fitting model* ke *neural network* adalah sebagai berikut. Pertama, pada *input layer* terdapat 3 *filter* dengan besar *input* sebesar 64 x 64. Ukuran *kernel filter* adalah 3,3. Selanjutnya model *neural network* dimasukkan ke dalam pemrograman yang kemudian dilanjutkan dengan *average pooling* dengan *dropout* sebesar 0,5. Kemudian, ada *fully connected layer* yang dilanjutkan dengan *multi-class classification* dengan menggunakan *optimizers* dengan *learning rate* sebesar 0,002.

Tabel 2 Parameter untuk Augmentasi

Metode	Default	Penyesuaian
Rotasi	0	40
Width shift	0	0,2
Heigh shift	0	0,2
Zoom range	-	0,2
Shear range	0	0,2
Horizontal flip	None	True

1.3. ResNet-50

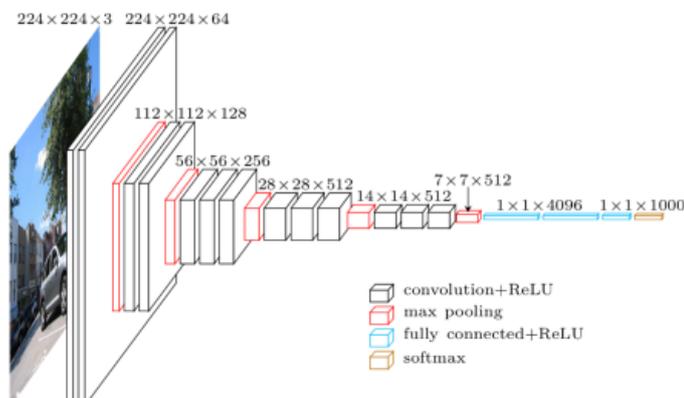
ResNet-50 merupakan variasi dari arsitektur ResNet yang memiliki 50 lapis yang sudah dilatih setidaknya pada 1 juta gambar yang ada pada database ImageNet.⁶ ResNet -50 terdiri dari lima tahapan yang di masing-masing tahapannya terdapat konvolusi dan *identity block*. Setiap blok konvolusi terdiri dari dua lapis konvolusi dan masing-masing *identity block* juga memiliki tiga lapis konvolusi. ResNet 50 sudah memiliki lebih dari 23 juta parameter yang dapat dilatih. Gambar 4 menunjukkan arsitektur resnet-50 yang terdiri dari lima tahap proses konvolusi tersebut.



Gambar 4 Arsitektur ResNet-50.¹³

1.4. VGG-16

VGG-16 merupakan arsitektur CNN yang memenangkan kompetisi ImageNet pada tahun 2014. VGG-16 dinilai sebagai salah satu model arsitektur yang baik sampai saat ini. VGG-16 ini tidak memiliki angka *hyperparameter* yang besar, namun VGG-16 ini hanya fokus dalam filter lapis konvolusi 3x3 dengan *stride* 1 dan selalu menggunakan *padding* yang sama dan filter *maxpool layer* 2x2 dengan *stride* 2. Di akhir arsitekturnya, VGG-16 ini memiliki 2 *fully connected layer* yang diikuti dengan *softmax* untuk *output*. Angka 16 dari VGG-16 ini menunjukkan bahwa VGG-16 memiliki 16 lapis dengan beban yang sama. Jaringan ini merupakan jaringan yang cukup besar dan memiliki sekitar 138 juta parameter. Gambar 5 menunjukkan arsitektur VGG-16 tersebut.



Gambar 5 Arsitektur dari VGG-16.⁸

1.5. Performance Evaluation Metrics

Mengacu pada beberapa penelitian terkait klasifikasi data pada penelitian sebelumnya,¹⁴⁻²⁰ metrik evaluasi menjadi parameter yang penting untuk mengukur performa dari model yang dihasilkan secara kuantitatif. Beberapa metrik evaluasi digunakan untuk membandingkan performa dari kedua model yaitu VGG-16 dan Resnet-50 dalam mengklasifikasikan citra *x-ray* ke dalam empat kelas. Penelitian ini dilakukan sebanyak 3 kali dengan menggunakan epoch yang berbeda yaitu 10, 200 dan 400 dengan *batch size* 64. Akurasi, presisi, sensitivitas, *F1-Score*, dan spesifisitas hasil penelitian dapat dihitung dengan menggunakan persamaan berikut.¹⁴⁻²⁰

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Sensitivitas} = \frac{TP}{TP+FN}$$

$$F1 = 2 \times \frac{\text{presisi} \times \text{sensitivitas}}{\text{presisi} + \text{sensitivitas}}$$

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FN+FP}$$

$$\text{Spesifisitas} = \frac{TN}{TN+FN}$$

TP (*true positive*) merupakan klasifikasi yang benar dari sebuah kelas positif dan TN (*true negative*) adalah klasifikasi yang benar pada kelas negatif. FP (*false positive*) adalah prediksi yang salah dari kelas positif dan FN (*false negative*) merupakan prediksi yang salah pada kelas negatif.⁹

1.6. Prediksi (*testing*)

Testing ini dilakukan untuk mengetahui apakah *training* yang dilakukan sudah berhasil dalam mengklasifikasikan sebuah gambar atau tidak. *Testing* ini juga dilakukan untuk mengetahui kecepatan model tersebut dapat mengklasifikasikan sebuah gambar. Hasil kecepatan yang didapatkan akan dibandingkan dengan menggunakan metode analisis *paired t-test* dengan level signifikansi sebesar 0,05.

Hasil

Hasil ini berdasarkan perbedaan dari kedua pendekatan yang berbeda, pertama dilakukan penelitian menggunakan VGG-16, kemudian dilanjutkan dengan penelitian menggunakan ResNet-50. Tabel 3 menunjukkan pengaruh perbedaan epoch terhadap tingkat akurasi model VGG-16.

Gambar 6 (a) menunjukkan grafik akurasi dan (b) grafik *loss* terhadap epochs yang berbeda. Untuk grafik akurasi epochs menunjukkan hasil yang baik dimana nilai akurasi terus

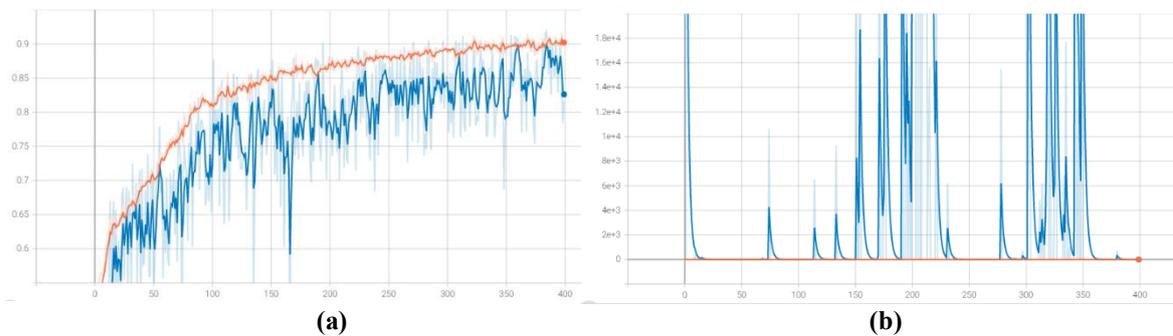
Research Article

mengalami kenaikan, namun pada saat mencapai epochs 200, kenaikan dari akurasi tidak lagi mengalami kenaikan yang signifikan..

Selanjutnya untuk pengaruh jumlah epoch terhadap tingkat akurasi model Resnet-50 dapat dilihat pada tabel 4. Tabel 4 menunjukkan bahwa saat menggunakan nilai epochs lebih besar, maka hasil akurasi yang didapatkan juga semakin besar. Pada saat nilai epochs 400 dan dilakukan *testing*, hasil menunjukkan nama kelas dengan *confidence level* sebesar 36,55% saja dengan waktu sekitar 5 detik. Pada saat nilai epochs 10, model ResNet-50 tidak bisa mengklasifikasikan gambar dengan benar. Dari tabel 2 juga dapat dilihat bahwa perbedaan akurasi dari nilai epochs 10 dengan epochs 100 sangat berbeda jauh dengan hanya mengubah parameter epochs saja. Ini menunjukkan bahwa akurasi akan terus meningkat saat dilakukan *finetuning* parameter lain pada model.

Tabel 3 Pengaruh Epochs pada Akurasi Model VGG-16

Epochs	Akurasi <i>Training</i> (%)	Akurasi Validasi (%)
10	59,51%	46,39%
200	89,84%	80,60%
400	91,36%	81,69%



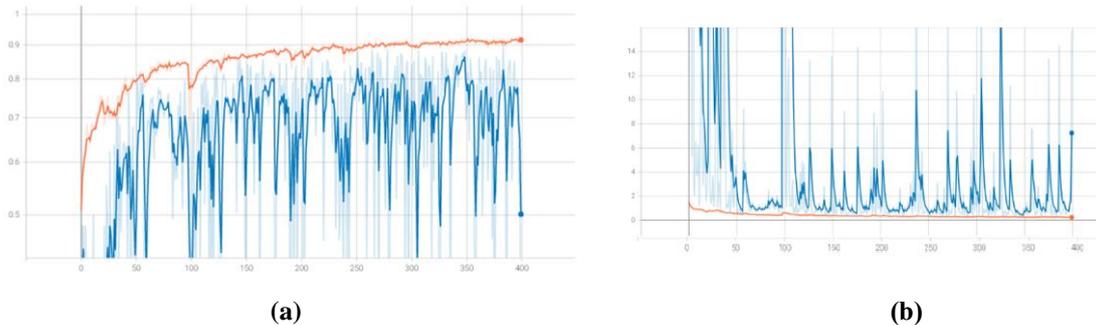
Gambar 6 (a) Akurasi-epochs 400, (b) Loss- epochs 400 dari VGG-16

Tabel 4 Pengaruh Epochs pada Akurasi Model ResNet-50

Epochs	Akurasi <i>Training</i> (%)	Akurasi Validasi (%)
10	67,46%	17,5%
200	88,4%	63,18%
400	90,25%	27,88%

Research Article

Gambar 7 menunjukkan grafik tingkat akurasi dan *loss* pada *epoch* yang berbeda dari model Resnet-50. Grafik tersebut menunjukkan bahwa nilai akurasi dan *loss* pada model Resnet-50 cenderung fluktuatif. Hal ini berbeda dengan VGG-16 pada gambar 6 yang cenderung meningkat secara stabil. Pada ResNet-50 yang dinilai memiliki fungsi yang optimal adalah dengan menggunakan epochs sekitar 340.



Gambar 7 (a) Akurasi-epochs 400 (b) Loss-epochs 400 dari Resnet -50

Pada penelitian ini, hasil dianalisis dengan menggunakan metrik evaluasi yang terdiri dari akurasi, spesifisitas, presisi, sensitivitas dan *F1-score*. Selain tingkat akurasi, kecepatan tes dari masing-masing model juga dievaluasi. Tabel 5 menunjukkan metrik evaluasi hasil perbandingan antar VGG-16 dan Resnet-50.

Dengan menggunakan analisis *paired t-test* yang langsung membandingkan antara masing-masing akurasi, spesivisitas, presisi, sensitivitas dan *F1-score*, tidak terdapat perbedaan yang signifikan antara VGG-16 dan Resnet-50 dimana *p value* = 0,4176, *t* = 0,9 (dalam hal ini h_0 = ResNet-50 dan VGG dianggap memiliki akurasi yang sama) dalam mengklasifikasikan citra medis *x-ray*.

Tabel 5 Metrik Evaluasi VGG-16 dan ResNet-50

Metrics	VGG-16	ResNet-50
Akurasi (%)	91,45	95,13
Spesifisitas (%)	96,89	96,44
Presisi (%)	92,58	92,85
Sensitivitas (%)	90,35	85,01
<i>F1-score</i> (%)	95,62	88,76
Speed Test (ms)*	19500	20400

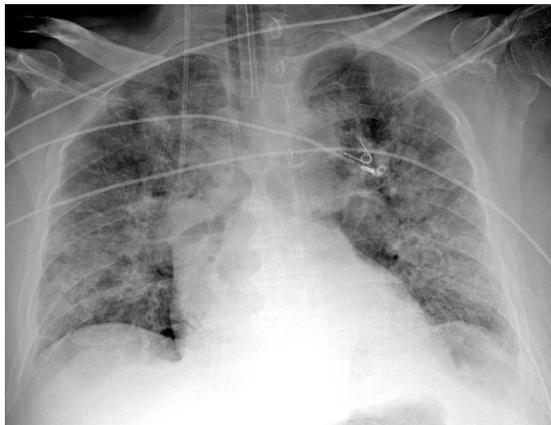
* *Speed test* = Rata-rata kecepatan dengan 30 kali percobaan

Research Article

Dari sisi kecepatan, VGG-16 memiliki waktu pemrosesan atau waktu inferensi yang lebih singkat dibandingkan model ResNet-50 yang mana diperoleh, $t = 0,09$, $p \text{ value} = 0,9289$ (dalam hal ini $h_0 = \text{ResNet-50}$ dan VGG dianggap memiliki waktu inferensi sama). Menurut hasil yang diperoleh pada penelitian ini, disimpulkan bahwa baik Resnet-50 dan VGG-16 berhasil mendapatkan nilai lebih dari 85% untuk semua nilai metrik evaluasi, namun dalam hal kecepatan, VGG-16 merupakan model yang lebih baik digunakan untuk mengklasifikasikan citra medis *x-ray*. Penelitian ini ditujukan untuk menentukan klasifikasi pada *data set x-ray* yang tersedia. Parameter yang digunakan sederhana sehingga mudah diimplementasikan dalam aplikasi nyata.

Pada tulisan ini telah dibahas perbandingan hasil akurasi untuk klasifikasi dari gambar *x-ray* dengan menggunakan ResNet50 dan VGG-16 dengan menggunakan Google Colab. Klasifikasi ini bertujuan untuk mengklasifikasikan citra medis ke dalam 4 kelas yaitu *x-ray* normal, *x-ray viral pneumonia*, *x-ray lung opacity* dan *x-ray Covid-19*, untuk membantu sistem pengambilan keputusan dalam diagnosis citra *x-ray* pasien terduga Covid-19. Dalam tulisan ini digunakan model *pre-trained* untuk dapat mengklasifikasikan citra medis secara lebih efisien.

Hasil yang diperoleh menunjukkan saat epochs tertentu yaitu 200, akurasi dari VGG-16 lebih besar dibandingkan ResNet-50. Proses *fine tuning* masih harus terus dilakukan untuk mendapatkan angka akurasi yang paling maksimal apabila ke depannya akan diterapkan pada *data set* yang baru. Pada penelitian ini juga terdapat keterbatasan, yang mana terdapat beberapa contoh data mengalami salah klasifikasi. Contohnya pada Gambar 8, seharusnya gambar tersebut masuk kategori normal, namun terprediksi sebagai gambar untuk kelas Covid. Hal tersebut dapat terjadi karena pada citra *x-ray* normal tersebut terdapat sedikit bercak putih, sehingga mengakibatkan terjadinya kesalahan prediksi tersebut.



Gambar 8 Contoh *False Positive*

Research Article

Hasil yang didapatkan menunjukkan bahwa model CNN VGG-16 merupakan model yang hasilnya lebih optimal dibandingkan dengan ResNet-50 dari sisi kecepatan, namun tidak dari sisi metrik evaluasi lainnya. Hasil analisis ini dibuktikan dengan menggunakan *paired t-test* yang menunjukkan bahwa tidak ada kenaikan akurasi, presisi, sensitivitas, spesifisitas dan *F1 score* saat menggunakan model ResNet-50. Hasil yang didapatkan ini dapat dipengaruhi oleh beberapa faktor seperti faktor perbedaan jumlah dan proporsi *data set* yang digunakan untuk melakukan *training*. Penelitian ini selanjutnya bisa terus dikembangkan dengan menambahkan data klinis aktual yang lebih banyak dengan mempertimbangkan berbagai jenis kondisi variasi kelainan paru-paru pada pasien, sehingga model yang dihasilkan lebih akurat dan salah klasifikasi seperti pada gambar 8 bisa diminimalisasi.

Simpulan

Berdasarkan hasil yang didapatkan, dapat disimpulkan bahwa VGG-16 memiliki kemampuan klasifikasi yang lebih baik dibandingkan ResNet-50 dalam klasifikasi data *X-ray* ini. Kemampuan ini sangat ditentukan *data set* dan parameter yang digunakan. Tidak terdapat perbedaan yang signifikan antara VGG-16 dan Resnet-50 dalam akurasi, spesivisitas, presisi, sensitivitas dan *F1-score*

Penelitian ini memiliki potensi yang baik dalam pengembangan radiologi lebih lanjut dan dapat digunakan sebagai sistem pendukung bagi para tenaga medis dalam mengambil keputusan. Penelitian lebih lanjut perlu dilakukan untuk meningkatkan akurasi. Aplikasi dari penelitian ini diharapkan dapat digunakan untuk membantu proses diagnosis pasien dengan tetap mendapatkan pengawasan penuh dari pihak pengambil keputusan.

Daftar Pustaka

1. Kim H, Goo JM, Kim TJ, Kim HY, Gu G, Gil B, et al. Effectiveness of radiologist training in improving reader agreement for Lung-RADS 4X categorization. *Eur Radiol*. 2021. [Cited 2022 August 2]. Available from: <https://doi.org/10.1007/s00330-021-07990-y>.
2. Barlow WE, Chen Chi, Patricia A. Carney, Stephen H. Taplin, Carl D'Orsi, Gary Cutter, et al. Accuracy of Screening Mammography Interpretation by Characteristics of Radiologists, JNCI: Journal of the National Cancer Institute. 2014; 96 (24): 1840–1850.
3. Müller D, Kramer F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Medical Imaging*. 2021;21(1). [Cited 2022 August 5]. Available from: <https://doi.org/10.1186/s12880-020-00543-7>.
4. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998;86(11):2278-324. doi: 10.1109/5.726791.
5. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science, vol 8689. Springer, Cham. 2014.
6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90. doi: 10.1145/3065386.
7. Victor Ikechukwu A, Murali S, Deepu R, Shivamurthy RC. ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global*

Research Article

- Transitions Proceedings. 2021;2(2). doi: 10.1016/j.glopro.2021.100036.
8. Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers*. 2019;11(9):1235. doi: 10.3390/cancers11091235.
 9. Rajpal S, Lakhyani N, Singh AK, Kohli R, Kumar N. Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images. *Chaos Solitons Fractals*. 2021;145:110749. doi: 10.1016/j.chaos.2021.110749.
 10. Ismael AM, Şengür A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications*. 2021;164:114054. doi: 10.1016/j.eswa.2021.114054.
 11. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahub ZB, et al. Can AI help in screening Viral and COVID-19 pneumonia? *IEEE Access*. 2020;8:132665-132676. [Cited 2022 January 19]. Available from: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>. doi: 10.1109/ACCESS.2020.3014131.
 12. Joseph Paul Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv:2003.11597, 2020. [Cited 2022 January 19]. Available from: <https://github.com/ieee8023/covid-chestxray-dataset>, <https://www.kaggle.com/pranavraikokte/covid19-image-dataset>.
 13. Dwivedi P. Understanding and Coding a ResNet in Keras-Towards Data Science. Medium. Towards Data Science; 2019 [Cited 2022 January 20]. Available from: <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>.
 14. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. 2022;15:210. doi: 10.1186/s13104-022-06096-y.
 15. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *medRxiv*. 2021. [Cited 2022 April 18]. Available from: <https://doi.org/10.1101/2021.04.07.21254975>.
 16. Ham HS, Lee HS, Chae JW, Cho HC, Cho HC. Improvement of Gastroscopy Classification Performance Through Image Augmentation Using a Gradient-Weighted Class Activation Map. *IEEE Access*. 2022;10:99361-99369. doi: 10.1109/ACCESS.2022.3207839.
 17. Junayed MS, Islam MB, Jeny AA, Sadeghzadeh A, Biswas T, Shah AFMS. ScarNet: Development and Validation of a Novel Deep CNN Model for Acne Scar Classification With a New Dataset. *IEEE Access*. 2022;10:1245-1258. doi: 10.1109/ACCESS.2021.3138021.
 18. Tang GS, Chow LS, Solihin MI, Ramli N, Gowdh NF, Rahmat K. Detection of COVID-19 Using Deep Convolutional Neural Network on Chest X-ray (CXR) Images. 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), ON, Canada, 2021, pp. 1-6. doi: 10.1109/CCECE53047.2021.9569064.
 19. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12:5979. doi: 10.1038/s41598-022-09954-8.
 20. Sait T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*. 2015;10(3):e0118432. [Cited 2022 April 18]. Available from: <https://doi.org/10.1371/journal.pone.0118432>.