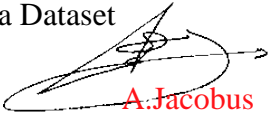


Data Balancing Methods on Radiographic Image Classification on Unbalance Dataset

Perbandingan Metode Penyeimbangan Data pada Klasifikasi Citra Radiografi pada Dataset Tidak Seimbang

Joshua A. Wuisan, Agustinus Jacobus, Sherwin R. U. A. Sompie
Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia
E-mails: axelwuisan@gmail.com, a.jacobus@unsrat.ac.id, aldo@unsrat.ac.id

Received: [date]; revised: [date]; accepted: [date]



A. Jacobus

Abstract — Covid-19 is a disease caused by a corona virus infection that infects the victim's respiratory tract. Covid-19 disease has a high infectious ability and if treated too late can result in death. Covid-19 has been a problem faced by everyone in the world since the end of 2019. Fast and accurate detection can save many lives. This study aims to develop a predictive model of COVID-19 detection based on radiographic images using a machine learning model from 4 categories of health status, namely positive covid, normal, lung opacity sufferers, and viral pneumonia sufferers. Deep learning is based on ResNet50 and MobileNetV2 and trials of undersampling and oversampling data balancing methods, and uses a confusion matrix for the evaluation process of model results. The model with the highest performance achieves 95.58% accuracy in the multi-class classification. Also based on the findings, we provide results from using a different data balancing approach or not using one at all.

Key words — classification, data balancing, deep neural network, machine learning, radiographic imagery.

Abstrak — Covid-19 adalah penyakit yang disebabkan oleh infeksi virus corona yang menginfeksi saluran pernapasan korban. Penyakit Covid-19 memiliki kemampuan menular yang tinggi dan jika terlambat ditangani dapat mengakibatkan kematian. Covid-19 telah menjadi masalah yang dihadapi oleh semua orang di dunia sejak akhir tahun 2019. Deteksi yang cepat dan akurat dapat menyelamatkan banyak nyawa. Penelitian ini bertujuan untuk mengembangkan model prediktif deteksi COVID-19 berdasarkan citra radiografi dengan menggunakan model pembelajaran mesin dari 4 kategori status kesehatan yaitu positif covid, normal, penderita *lung opacity*, dan penderita *viral pneumonia*. *Deep learning* didasarkan pada ResNet50 dan MobileNetV2 serta uji coba metode penyeimbangan data *undersampling* dan *oversampling*, dan menggunakan *confusion matrix* untuk proses evaluasi hasil model. Model dengan performa tertinggi mencapai akurasi 95,58% pada klasifikasi multi-kelas. Juga berdasarkan temuan, kami memberikan hasil dari menggunakan pendekatan penyeimbangan data yang berbeda atau tidak menggunakan sama sekali.

Kata kunci — citra radiografi, *convolutional neural network*, klasifikasi, *machine learning*, penyeimbangan data.

I. PENDAHULUAN

Pandemik COVID-19 yang sudah berlangsung dari penghujung tahun 2019 masih melanda dunia. Penyakit COVID-19 disebabkan oleh infeksi virus corona yang sangat

menular. COVID-19 diduga berawal dari hewan yang terinfeksi di pasar lokal kota Wuhan, China kemudian virus tersebut bertransmisi ke manusia. Penularan bisa melalui kontak fisik dengan hewan yang terinfeksi kemudian korban ada juga kontak fisik dengan orang lain yang menyebabkan virus ini bisa menyebar.[1]

Serangan virus sering ditemukan menyerang paru-paru korban karena virus masuk dari rongga pernafasan. Berdasarkan laman <https://www.worldometers.info/coronavirus/> yang diakses pada tanggal 3 Desember 2021, total sejumlah 264,439,993 kasus penderita COVID-19 di seluruh dunia dan 5,249,487 orang meninggal karena COVID-19. Prosedur standar yang dilakukan untuk melakukan pendeteksian virus adalah diagnosis melalui uji *Reverse Transcription – polymerase chain reaction* (RT-PCR) untuk memberikan diagnosa yang akurat.

Beberapa penelitian sebelumnya sudah melakukan uji coba terhadap metode pendeteksian virus COVID-19 menggunakan *deep learning* pada klasifikasi citra x-ray dada pasien. Pendekatan *deep learning* membuat analisis citra radiografi menjadi otomatis dan dapat mengekstraksi informasi berguna dari hasil x-ray dada.

Namun demikian, untuk membangun model *deep learning* yang baik, maka diperlukan data yang baik juga. Kelemahan dari pengembangan *deep learning* sejauh ini adalah ketersediaan data yang minim atau data yang tersedia tidak memiliki jumlah yang seimbang. Ini bisa berakibat pada hasil *deep learning* yang tidak akurat dan tidak mencapai performa terbaik.

Penelitian ini bertujuan untuk mencoba metode penyeimbangan data pada *dataset* citra radiografi yang tidak seimbang. Pengembangan model akan dilakukan dengan pendekatan *transfer learning*.

A. Penelitian Terkait

Penelitian yang dilakukan oleh Basu melakukan *transfer learning* pada 3 *pretrained model* dengan menggunakan 2 *dataset* yang berbeda. *Dataset A* terdiri dari 2 kategori yaitu kategori normal sebanyak 47000 lebih data gambar dan kategori berpenyakit sebanyak 41000 lebih data gambar. Gambar pada kategori berpenyakit mencakup penyakit paru-paru, jantung, dan penyakit yang berhubungan dengan dada

lainnya. Untuk *dataset B* terdiri atas 4 kategori yaitu normal (350 gambar), *pneumonia* (322 gambar), penyakit lainya (300 gambar), dan covid (305 gambar). Setelah model dilatih dengan *dataset A*, *transfer learning* dilakukan terhadap *dataset B*. *Pretrained model* dan hasil dari pelatihannya adalah AlexNet (82,98%), VGGNet (90,13%), dan ResNet (85,98%).[2]

Penelitian yang dilakukan oleh Hammoudi melakukan penyesuaian kepada beberapa *pretrained model* seperti ResNet34, ResNet50, DenseNet169, VGG-19, Inception ResNetV2 dan RNN. *Dataset* yang digunakan adalah *Chest X-ray Images (Pneumonia) dataset* yang terdiri atas 3 kategori yaitu, normal, *bacterial pneumonia*, dan *viral pneumonia*. Pembagian datanya merata sebanyak 1345 data latih dan 148 uji untuk setiap kategori. Sedangkan untuk pendeteksian covid, Hammoudi melakukan *blind test* pada 145 gambar x-ray positif covid dengan asumsi bahwa covid diklasifikasikan sebagai *viral pneumonia*. Hasilnya adalah DenseNet169 memiliki akurasi terbaik sebesar 95,72%. [3]

Penelitian yang dilakukan oleh Hariyani melakukan transfer learning menggunakan ResNet50 pada klasifikasi dua kategori, normal dan covid. Dalam penelitian ini dikembangkan 3 model yang berbeda dari Sebagian arsitektur ResNet50 yang dianggap memiliki *high level feature* dan cocok untuk dilatih kembali dengan data gambar x-ray. Jumlah data yang digunakan adalah 42 gambar untuk kategori covid dan 234 gambar kategori normal. Salah satu model yang dibuat mencapai akurasi 99%. [4]

Penelitian yang dilakukan oleh Farooq mengusulkan sebuah metode untuk membangun *convolutional neural network* dengan arsitektur yang memiliki jumlah parameter yang lebih sedikit untuk mendapatkan keseimbangan antara performa dan kemampuan komputasional. Proses pelatihan yang dilakukan diharapkan hanya memerlukan sedikit jumlah *epoch* sehingga mempercepat waktu pelatihan. *Dataset* yang digunakan adalah COVIDx *dataset* dengan 4 kategori yaitu, normal (1203 gambar), *bacterial pneumonia* (931 gambar), *viral pneumonia* (660 gambar), dan covid (98 gambar). Untuk menangani masalah ketidak seimbangan data, penerapan augmentasi data dilakukan. Proses pelatihan dilakukan terhadap *pretrained model* ResNet50, dimana beberapa *layer* dari *pretrained model* digantikan dengan konfigurasi dari Farooq. Proses pelatihan dilakukan sebanyak 3 kali dengan ukuran gambar input yang diganti pada setiap proses. Pelatihan pertama dilakukan dengan gambar 128x128 pixel kemudian model dilatih kembali dengan ukuran gambar 224x224pixel dan terakhir dengan gambar 299x299pixel. Hasilnya adalah sebuah model yang mereka sebut COVID-ResNet dengan akurasi 96.23%, namun Farooq menyatakan model ini masih perlu pelatihan kembali dengan data yang lebih banyak sebelum diaplikasikan pada bidang medis. [5]

Penelitian yang dilakukan oleh Shelke membangun 3 model dengan tugas klasifikasi yang berbeda-beda untuk tujuan otomatis *screening* COVID-19. Penelitian yang dilakukan adalah penggunaan *transfer learning* model VGG-16 yang digunakan untuk mengklasifikasikan gambar x-ray terhadap kategori normal, tuberkulosis, atau *pneumonia*. Hasil klasifikasi dari VGG-16 yang diprediksi sebagai *pneumonia* akan menjadi input pada model DenseNet-161 untuk

diklasifikasikan sebagai covid atau *pneumonia* lainnya. Jika hasil prediksi dari model DenseNet-161 adalah covid, maka gambar tersebut akan menjadi input bagi model ResNet18 untuk diklasifikasikan dalam kategori covid parah, sedang, atau tidak parah. Model VGG-16 dilatih dengan 388 gambar kategori normal, 500 gambar kategori pneumonia yang didalamnya sudah termasuk kategori pneumonia lainnya dan COVID-19, dan 303 gambar kategori tuberkulosis. Untuk pelatihan model DenseNet-161 data pelatihannya sebanyak 500 gambar untuk kategori pneumonia dan 500 gambar untuk kategori COVID-19. Dan untuk pelatihan ResNet18 digunakan 80 gambar tanpa dijelaskan pembagian per kategorinya. Hasilnya adalah model VGG-16 dengan akurasi rata-rata 95,9%, DenseNet-161 98,9% dan ResNet18 76% dari bagian pengklasifikasian masing-masing. [6]

Berdasarkan penelitian-penelitian terkait yang telah dilakukan sebelumnya, sebagian besar penelitian terdapat kelemahan pada jumlah data yang tidak seimbang dan jumlah data gambar per kategorinya yang sedikit. *Deep learning* sensitif dengan data yang digunakan, jika data terlalu sedikit bisa menghasilkan performa yang tidak konsisten dan jika pelatihan dilakukan dengan *dataset* yang tidak seimbang, model bisa mengalami *overfitting* pada kategori mayoritas.

B. Artificial Intelligence

Artificial intelligence adalah keilmuan dan rekayasa dalam membuat komputer memiliki kemampuan mengumpulkan Informasi dan memproses kecerdasan (menyusun dan menafsirkan) serta menyajikan hasil proses penafsiran tersebut ke dalam bentuk yang dimengerti oleh manusia. [7]

C. Machine Learning

Machine learning adalah cabang keilmuan dari *artificial intelligence* yang merupakan pengembangan algoritma yang membuat komputer bisa belajar dari data yang ada tanpa pemrograman eksplisit. *Machine learning* akan belajar dan melakukan tugasnya tanpa ada instruksi dari penggunaannya. [8]

D. Deep Learning

Deep learning adalah metode pembelajaran mesin yang diawasi dengan struktur berupa tiruan dari jaringan saraf. *Deep learning* biasanya diimplementasikan sebagai salah satu bentuk *supervised learning* dengan keunggulan skalabilitasnya, dimana semakin besar sebuah *deep learning* maka semakin besar pula kemampuan ekstraksi fiturnya. [9]

E. Convolutional Neural Network

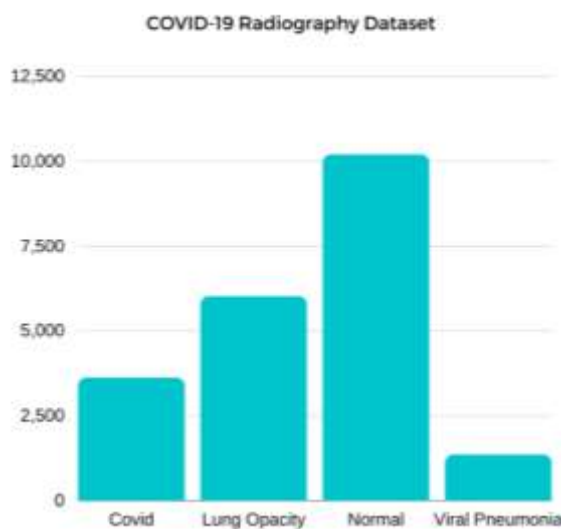
Convolutional neural network adalah salah satu arsitektur dari *deep learning* berupa representasi *neural network* dengan lapisan yang banyak. Dengan struktur yang dalam ini membuat model *convolutional neural network* bisa mendapatkan representasi karakteristik dari data yang disajikan melalui hasil transformasi non linier dan perkiraan fungsi non linier. Struktur *convolutional neural network* biasanya terdiri atas ekstraktor fitur yang didalamnya terdapat

beberapa lapisan *convolutional* dan biasanya diikuti dengan lapisan *pooling* serta sebuah pengklasifikasi *softmax*. Tugas dari lapisan *convolutional* adalah ekstraksi fitur, sedangkan lapisan *pooling* bertugas mengurangi dimensi data untuk menghemat waktu komputasi. Hasil ekstraksi fitur yang didapatkan kemudian masuk ke lapisan *softmax* untuk pengklasifikasian.[10]

F. Tensorflow

Tensorflow adalah sebuah *open-source framework* yang dikembangkan oleh *Google researchers* dengan kemampuan untuk beroperasi pada skala besar di dalam lingkungan heterogen. *Tensorflow* bisa melakukan komputasi antara banyak mesin dalam sebuah kelompok, penggunaan CPU dengan beberapa *core*, GPU, dan juga perangkat komputasi khusus yang dikenal sebagai *Tensor Processing Units* (TPUs). Arsitektur ini memberikan fleksibilitas kepada pengembang untuk bereksperimen. *Tensorflow* bisa mendukung beberapa aplikasi, dengan fokus pada pelatihan dan inferensi untuk *deep learning*. [11] Tujuan dari *tensorflow* adalah untuk melakukan *machine learning*, *deep learning*, serta kegiatan analisis dan prediktif lainnya. *Tensorflow* menggabungkan aljabar komputasi dengan teknik optimasi kompilasi untuk memfasilitasi perhitungan ekspresi matematika. Fitur yang terdapat dalam *tensorflow* adalah:

1. Mendefinisikan, mengoptimalkan, dan menghitung secara matematis dengan melibatkan *array multidimension* (*tensors*).
2. Sebuah *library* pendukung *deep learning* dan teknik *machine learning*.
3. Pemakaian GPU, mampu menjalankan kode yang sama dan menjalankannya di CPU atau GPU, mengetahui bagian mana yang harus dipindahkan ke GPU, manajemen dan optimalisasi memori terhadap data yang digunakan.
4. Skalabilitas komputasi yang tinggi pada keseluruhan mesin terhadap kumpulan data yang besar.



Gambar 1. Grafik Perbandingan Jumlah Data antar Kategori

II. METODE

A. COVID-19 Radiography Dataset

Penelitian ini dilakukan dengan menggunakan sebuah *dataset* publik yang bisa diakses pada website Kaggle. *Dataset COVID-19 Radiography Dataset* merupakan hasil pengumpulan data gambar radiografi x-ray dada dari beberapa sumber ke dalam satu *dataset* oleh Qatar University, University of Dhaka dan kolaborasi bersama dokter medis.[12], [13] Ukuran data gambarnya adalah 299x299 pixel, terdiri atas 4 kategori yaitu 1. Normal (tidak ada infeksi), 2. *Lung Opacity*, 3. *Viral Pneumonia*, 4. Positif Covid.

Pada versi dari *dataset* yang sedang digunakan terdapat 10192 data untuk kategori normal, 6012 data untuk kategori *lung opacity*, 1345 data untuk kategori *viral pneumonia*, dan 3616 data untuk kategori positif covid. Bisa dilihat bahwa ada perbedaan jumlah data antar kategori dan jumlah kategori covid bukanlah kategori dengan jumlah terdominan.

Gambar 1 menunjukkan grafik perbandingan jumlah data. *Dataset* kemudian dibagi menjadi 3 bagian yaitu data latih sebesar 80%, data uji sebesar 20%, dan 20% dari data latih dijadikan sebagai data validasi. Tabel I menunjukkan jumlah data dari setiap partisi setelah *dataset* dibagi.

B. Sampling

Gambar 1 menunjukkan adanya ketidakseimbangan jumlah data antar kategori. *Sampling* dilakukan pada data latih dengan tujuan untuk menyeimbangkan jumlah data dari 4 kategori yang ada agar hasil pelatihan model tidak condong ke kategori yang memiliki jumlah data terbanyak. Dalam penelitian ini digunakan 2 metode penyeimbangan data yaitu metode *undersampling* dan metode *oversampling*.

1) Undersampling

Metode menghilangkan/menghapus beberapa data pada kategori mayoritas.

2) Oversampling

Metode menduplikasi beberapa data pada kategori minoritas.

Sebagai tujuan pembandingan, maka akan ada 3 model yang dibuat, model A yang dilatih tanpa penyeimbangan data, model B yang dilatih dengan metode penyeimbangan data *undersampling*, dan model C yang dilatih dengan metode penyeimbangan data *oversampling*.

C. Data Preprocessing

Data *preprocessing* atau prapemrosesan data bertujuan untuk mempersiapkan data agar bisa digunakan sebagai data latih dan data uji terhadap model yang akan dibuat. Data *preprocessing* yang dilakukan dalam penelitian ini adalah sebagai berikut:

1) Resize data

Ukuran data gambar awal 299x299 pixel diubah menjadi 224x224 pixel.

2) Augmentasi data

Augmentasi data bertujuan untuk menambah variasi data dengan mengaplikasikan transformasi pada data gambar. Pada penelitian ini, transformasi yang digunakan adalah

horizontal flip saja. Data augmentasi hanya diterapkan pada data latih.

3) Preprocessing input function

Mengaplikasikan fungsi yang mengolah data sesuai dengan input *pretrained model* ketika dilatih pada *dataset* ImageNet.

D. Arsitektur Jaringan

Terdapat arsitektur *convolutional neural network* yang sudah dilatih sebelumnya dan menghasilkan kemampuan klasifikasi yang baik pada berbagai macam jenis pengklasifikasian. Pada penelitian ini digunakan *pretrained model* ResNet50 sebagai basis arsitektur dari model yang akan dibangun. Serta MobileNetV2 digunakan sebagai pembandingan jika hasil latih dengan ResNet50 tidak memuaskan.

Untuk tujuan *transfer learning*, kostum *fully connected layer* dibuat sesuai kebutuhan dan karakteristik dari hasil latih ketiga model.

E. Evaluasi

Precision, *recall*, dan *f1-score* adalah metrik yang digunakan untuk mengevaluasi model. Untuk mendapatkan nilai dari metrik tersebut, sebelumnya akan didefinisikan terlebih dahulu *true positive* (TP), *false positive* (FP), *false negative* (FN), dan *true negative* (TN) seperti yang ditunjukkan pada tabel II. TP didefinisikan sebagai data positif yang diprediksi sebagai kelas positif, TN didefinisikan sebagai data negatif yang diprediksi sebagai kelas negatif, sedangkan FP didefinisikan sebagai data negatif diprediksi sebagai kelas positif dan FN didefinisikan sebagai data positif diprediksi sebagai kelas negatif.

Nilai *precision* menggambarkan perbandingan TP terhadap total data yang diprediksi positif, nilai *recall* menggambarkan perbandingan TP terhadap total data positif, dan *f1-score* adalah perbandingan rata-rata antara *precision* dan *recall*.

TABEL I
HASIL JUMLAH DATA TIAP PARTISI SETELAH DATASET DIBAGI

Data	Jumlah
Latih	13545
Validasi	3387
Uji	4233

TABEL II
CONFUSION MATRIX

		Kategori sebenarnya	
		Positif	Negatif
Prediksi	Positif	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

TABEL III
PEMBAGIAN DATA LATIH SEBELUM SAMPLING

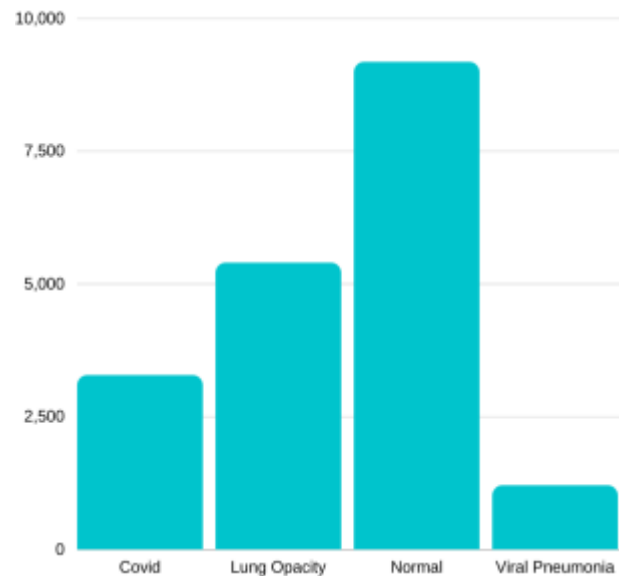
Metode Penyeimbangan Data	Data Latih	Covid	Lung Opacity	Normal	Viral Pneumonia
-	13545	2295	3909	6483	858
<i>Undersampling</i>	13545	2379	3825	6509	832
<i>Oversampling</i>	13545	2379	3825	6509	832

III. HASIL DAN PEMBAHASAN

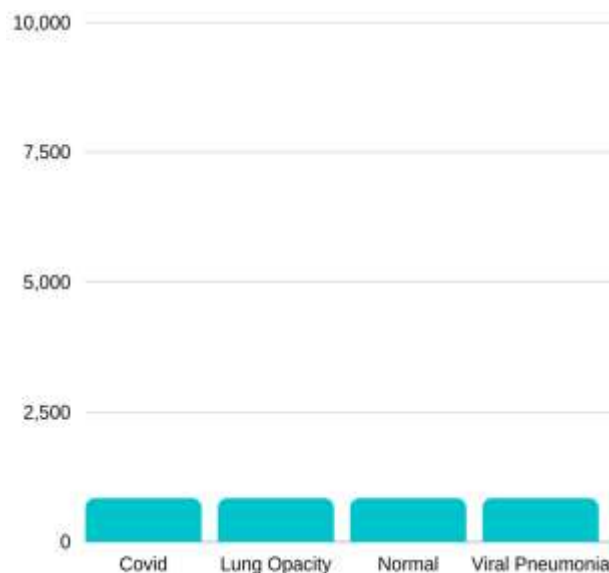
Bagian ini menjelaskan hasil *sampling* dari kedua metode penyeimbangan data serta hasil pelatihan model menggunakan data yang telah diseimbangkan jumlahnya. Dan sebagai pembandingan, model yang dilatih tanpa penyeimbangan data ditambahkan juga.

A. Hasil Sampling

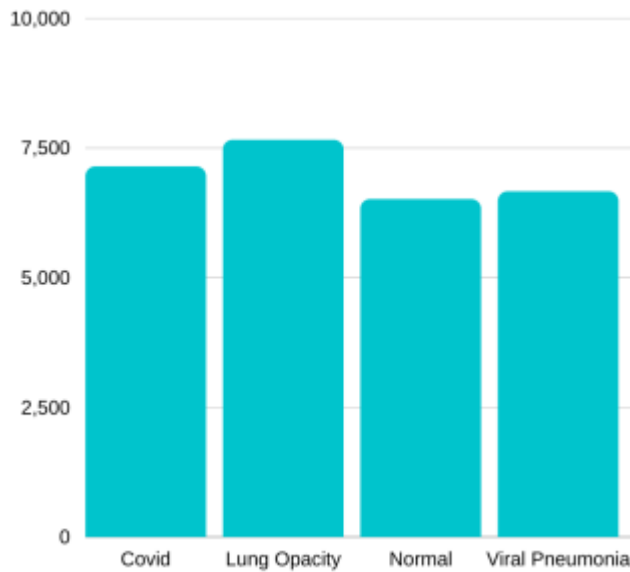
Sampling dilakukan pada data latih saja, perincian data latih per kategori sebelum dan sesudah *sampling* bisa dilihat pada tabel III dan tabel IV. Gambar 2 menunjukkan grafik perbandingan jumlah data latih sebelum *sampling*, sedangkan gambar 3 menunjukkan grafik perbandingan jumlah data latih setelah *undersampling* dan gambar 4 menunjukkan grafik perbandingan jumlah data latih setelah *oversampling*.



Gambar 2. Grafik Perbandingan Jumlah Data Latih Sebelum *Sampling*



Gambar 3. Grafik Perbandingan Jumlah Data Latih Sesudah *Undersampling*



Gambar 4. Grafik Perbandingan Jumlah Data Latih Sesudah *Oversampling*

TABEL IV
PEMBAGIAN DATA LATIH SESUDAH SAMPLING

Metode Penyeimbangan Data	Data Latih	Covid	Lung Opacity	Normal	Viral Pneumonia
-	19048	3278	5394	9176	1200
<i>Undersampling</i>	3328	832	832	832	832
<i>Oversampling</i>	27952	7137	7650	6509	6656

Data latih yang sudah diseimbangkan, data validasi, dan data uji diproses sesuai dengan bab II bagian C *Data Preprocessing*. Total data latih sesudah penyeimbangan data bisa dilihat pada tabel III. Tabel I menunjukkan data validasi berjumlah 3387 data, dan untuk data uji berjumlah 4233. Jumlah data validasi dan data uji tidak berubah karena penyeimbangan data hanya dilakukan pada data latih saja.

B. Pembuatan Model

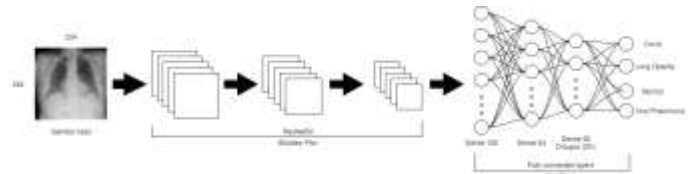
Tuning awal pelatihan dari ketiga model dibuat sama agar bisa dilihat hasil pelatihan masing-masing model dan proses *fine-tuning* model bisa dilakukan. *Tuning* awal tersebut merupakan pelatihan dengan basis arsitektur ResNet50 yang keseluruhan parameternya dilatih kembali terhadap data latih. *Batch size* yang digunakan sebesar 32 untuk semua model, pelatihan dilakukan sebanyak 20 iterasi (*epochs*), dengan ukuran gambar input sebesar 224x224 pixel, menggunakan *optimizer* Adamax, dan *learning rate* awal sebesar 0,001 Tabel V menunjukkan *tuning* pelatihan yang digunakan dan gambar 4 menunjukkan arsitektur *convolutional* model awal yang diterapkan pada ketiga model yang akan dikembangkan.

Proses pelatihan juga ditambahkan dengan 2 fungsi *callback* untuk membantu mempercepat dan menghasilkan model yang lebih baik. Fungsi tersebut adalah 1. *EarlyStopping* yang mengatur pemberhentian proses pelatihan jika setelah 3 iterasi pelatihan tidak terdapat perkembangan pada nilai *loss*, 2. *ReduceLROnPlateau* yang berfungsi untuk memperkecil nilai *learning rate* setelah 1 iterasi tanpa perkembangan pada nilai *loss*.

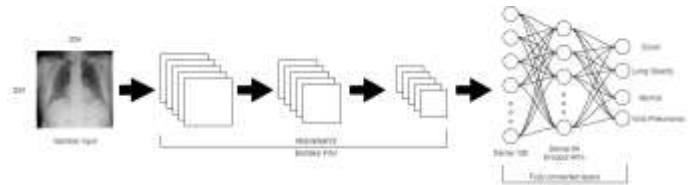
Hasil pengujian *tuning* awal dapat dilihat pada tabel VI. Dapat dilihat bahwa ketiga model berhasil mencapai akurasi

TABEL V
TUNING PELATIHAN AWAL UNTUK KETIGA MODEL

Basis Arsitektur	Batch size	Epochs	Input size	Input shape	Optimizer	Learning rate
Resnet50	32	20	224px	224, 224, 3	Adamax	0.001



Gambar 5. Arsitektur *Convolutional Network* Awal untuk Ketiga Model



Gambar 6. Arsitektur *Convolutional Network* Model B Setelah *Fine-tuning*

TABEL VI
HASIL PELATIHAN KETIGA MODEL DENGAN *TUNING* AWAL

Model	Loss	Akurasi
Model A	39	96
Model B	64	93
Model C	35	95

lebih dari 90%. Namun perlu diperhatikan pula nilai *loss*-nya. Model A yang dilatih tanpa penyeimbangan data memiliki nilai *loss* sebesar 39% dan model C yang dilatih dengan *oversampling* data memiliki nilai *loss* sebesar 35%, sedangkan model B yang dilatih dengan *undersampling* data memiliki nilai *loss* sebesar 64%.

Pengujian selanjutnya dilakukan terhadap model B guna menurunkan nilai *loss* yang akan dihasilkan. Pengujian untuk mengurangi *loss* model B adalah mengganti basis arsitektur dari ResNet50 menjadi MobileNetV2 dengan tujuan mengurangi jumlah parameter yang perlu untuk dilatih dari 20 juta lebih parameter pada ResNet50 menjadi 2 juta lebih pada MobileNetV2. Hasil dari pergantian basis arsitekturnya adalah penurunan nilai *loss* dari 67% menjadi 48%. Penurunan tersebut cukup besar namun masih berbeda jauh dengan model A dan model C maka dengan memperkecil *batch size* dari 32 menjadi 8, proses ini menambah jumlah data yang akan dipelajari oleh model pada setiap iterasinya serta mengurangi jumlah neuron pada *fully connected layer* model B. Gambar 5 menunjukkan arsitektur model B setelah dikurangi jumlah neuronnya. Hasil dari pengecilan *batch size* dan pengurangan jumlah neuron adalah model B memiliki nilai *loss* 35%.

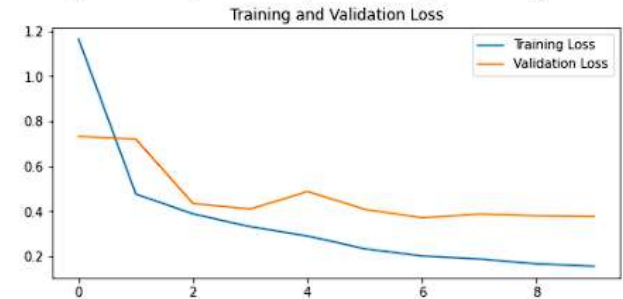
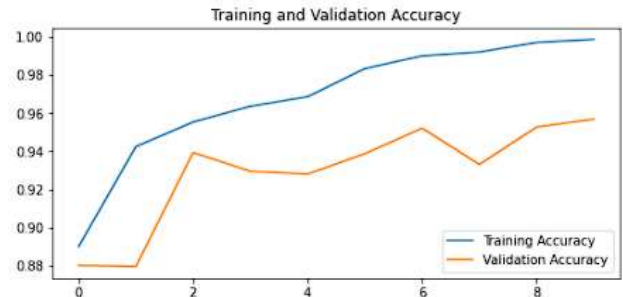
Dari hasil *fine-tuning* terhadap model B, maka *tuning* akhir dari model B dilatih dengan basis arsitektur MobileNetV2 dengan *batch size* 8. Ukuran gambar input, *optimizer*, dan *learning rate* awalnya tetap sama, *Tuning* akhir dari ketiga model dapat dilihat pada tabel VII dan hasil pelatihannya dapat dilihat pada tabel VIII.

TABEL VII
TUNING AKHIR PELATIHAN UNTUK KETIGA MODEL

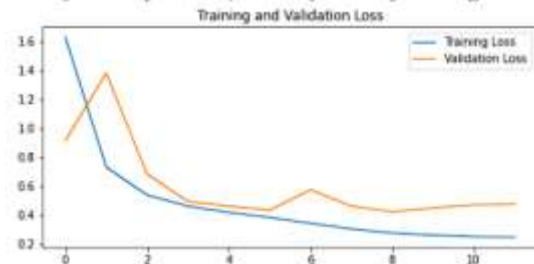
Basis Arsitektur	Batch size	Epochs	Input size	Input shape	Optimizer	Learning rate
Resnet50	32	20	224px	224, 3	Adamax	0.001
MobileNetV2	8	20	224px	224, 3	Adamax	0.001
Resnet50	32	20	224px	224, 3	Adamax	0.001

TABEL VIII
HASIL PELATIHAN KETIGA MODEL DENGAN TUNING AKHIR

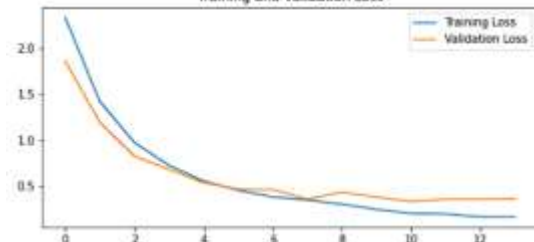
Model	Loss	Akurasi
Model A	39	96
Model B	35	94
Model C	35	95



Gambar 9. Grafik Perbandingan Hasil Latih dan Validasi Model C



Gambar 7. Grafik Perbandingan Hasil Latih dan Validasi Model A



Gambar 8. Grafik Perbandingan Hasil Latih dan Validasi Model B

TABEL IX
CONFUSION MATRIX MODEL A

Kategori	Covid	Lung Opacity	Normal	Viral Pneumonia
Covid	689	0	3	2
Lung Opacity	5	1152	68	0
Normal	1	92	1941	7
Viral Pneumonia	0	0	9	264

TABEL X
CONFUSION MATRIX MODEL B

Kategori	Covid	Lung Opacity	Normal	Viral Pneumonia
Covid	680	3	9	2
Lung Opacity	8	1050	129	1
Normal	13	92	1961	9
Viral Pneumonia	0	0	5	271

TABEL XI
CONFUSION MATRIX MODEL C

Kategori	Covid	Lung Opacity	Normal	Viral Pneumonia
Covid	673	3	15	3
Lung Opacity	1	1090	97	0
Normal	6	69	1987	13
Viral Pneumonia	0	0	1	275

TABEL XII
RECALL DARI KETIGA MODEL

Model	Covid	Lung Opacity	Normal	Viral Pneumonia
Model A	99	94	95	97

Model B	98	88	95	98
Model C	97	92	96	100

TABEL XII
PRECISION DARI KETIGA MODEL

Model	Covid	Lung Opacity	Normal	Viral Pneumonia
Model A	99	93	96	97
Model B	97	92	93	96
Model C	98	94	95	95

TABEL XIV
F1-SCORE DARI KETIGA MODEL

Model	Covid	Lung Opacity	Normal	Viral Pneumonia
Model A	99	93	96	97
Model B	97	90	94	97
Model C	98	93	95	97

Gambar 7, gambar 8, dan gambar 9 menunjukkan grafik perbandingan akurasi dan *loss* antara hasil latih dan hasil validasi.

C. Evaluasi Model

Setelah model dilatih, performa model akan diuji menggunakan data uji. Hasil prediksi akan dibandingkan menggunakan *confusion matrix*, tabel *confusion matrix* dari ketiga model bisa dilihat pada tabel IX, tabel X, dan tabel XI. Selanjutnya akan diukur performa dari model dengan melihat hasil sensitivitas (*recall*), *precision*, dan *f1-score*. Tabel XII, tabel XIII dan tabel XIV menunjukkan nilai *recall*, *precision* dan *f1-score* dari ketiga model.

Confusion matrix menunjukkan kategori covid dan *viral pneumonia* bisa diklasifikasikan dengan baik oleh ketiga model diikuti kategori normal, dan terakhir kategori *lung opacity*. Sebagian hasil klasifikasi dari normal diklasifikasikan sebagai *lung opacity* dan sebaliknya. Ini bisa menandakan model yang masih belum bisa membedakan antara kondisi normal dan *lung opacity* dengan akurat namun nilai *recall* dan *precision* masih rata-rata 90% yang menandakan performa model tidak terlalu buruk untuk kedua kategori tersebut. Sedangkan untuk kategori covid dan *viral pneumonia* sudah bisa diklasifikasikan dengan baik oleh ketiga model. Dengan catatan bahwa ada beberapa kasus covid yang diklasifikasikan normal dan sebaliknya. Jumlahnya kecil dibandingkan dengan kesalahan klasifikasi antara normal dan *lung opacity* namun tetap untuk perlu diperhatikan karena kesalahan klasifikasi antara kategori covid dan normal adalah masalah yang signifikan. Namun nilai *recall* dan *precision* kategori covid sangatlah tinggi menunjukkan performa model yang baik.

Model A memiliki jumlah kesalahan klasifikasi yang sangat kecil antara kategori covid dan normal, dimana hanya 1 gambar normal yang diklasifikasikan sebagai covid dan 3 gambar covid yang diklasifikasikan sebagai normal. Rendahnya kesalahan klasifikasi antara kedua kategori ini

menjadi penanda yang baik bahwa model A bisa membedakan antara kategori normal dan covid lebih baik, dibandingkan dengan kedua model lainnya.

Nilai *recall* dari ketiga model terhadap 4 kategori yang ada memberikan rata-rata nilai sebesar 90% yang berarti ketiga model berhasil melakukan prediksi dengan benar dari hampir keseluruhan data aktual pada keempat kategori tersebut dengan tambahan, bahwa jika dilihat pada tabel XII, nilai *recall* pada kategori covid dan *viral pneumonia* lebih tinggi daripada kategori *lung opacity* dan normal. Sedangkan untuk nilai *precision* hampir semua memberikan nilai 90% yang berarti model dapat menghasilkan prediksi yang sebagian besar benar terhadap kelas aktualnya, terkecuali nilai *recall* dan *precision* ketiga model terhadap kategori *lung opacity* yang lebih rendah dibandingkan terhadap kategori lainnya. Ini menunjukkan model masih bingung membedakan *lung opacity* diantara keempat kategori.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan, dapat ditarik kesimpulan bahwa model yang dilatih dengan metode penyeimbangan data *undersampling* dan *oversampling* menghasilkan nilai *loss* terendah, dan model yang dilatih tanpa penyeimbangan data menghasilkan nilai akurasi tertinggi. Hasil dari pelatihan dan pengujian ketiga model menunjukkan bahwa model tersebut dapat melakukan klasifikasi dengan baik. dengan model A memberikan hasil terbaik berdasarkan nilai *recall*, *precision* dan *f1-score* rata-rata dari semua model serta hasil pengujian menggunakan data latih menunjukkan model A minim melakukan kesalahan klasifikasi antara kategori normal dan covid. Metode penyeimbangan data latih yang dilakukan belum menunjukkan perbedaan hasil yang signifikan antara ketiga model.

B. Saran

Dalam penelitian ini masih terdapat hal yang dapat dikembangkan lagi, seperti menambahkan lebih banyak data ketika ada data baru yang tersedia, penggunaan atau pengembangan algoritma penyeimbangan data gambar yang lebih akurat. Pendekatan untuk menyeimbangkan keseluruhan dataset, bukan hanya menyeimbangkan data latihnya saja. Serta uji coba dengan basis arsitektur lainnya untuk melihat hasil yang paling efektif.

V. KUTIPAN

- [1] Z. W. Ye, S. Yuan, K. S. Yuen, S. Y. Fung, C. P. Chan, and D. Y. Jin, "Zoonotic origins of human coronaviruses,"
- [2] S. Basu, S. Mitra, and N. Saha, "Deep Learning for Screening COVID-19 using Chest X-Ray Images," in *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, Dec. 2020, pp. 2521–2527. doi: 10.1109/SSCI47803.2020.9308571.
- [3] K. Hammoudi *et al.*, "Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19," *Journal of Medical Systems*, vol. 45, no. 7, Jul. 2021, doi: 10.1007/s10916-021-01745-4.
- [4] Y. S. Hariyani, S. Hadiyoso, and T. S. Siadari, "Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network,"

- ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 8, no. 2, p. 443, May 2020, doi: 10.26760/elkomika.v8i2.443.
- [5] M. Farooq and A. Hafeez, “COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs,” 2020. [Online]. Available: <https://github.com/lindawang/COVID-Net>.
- [6] A. Shelke *et al.*, “Chest X-ray Classification Using Deep Learning for Automated COVID-19 Screening,” *SN Computer Science*, vol. 2, no. 4, Jul. 2021, doi: 10.1007/s42979-021-00695-5.
- [7] P. Dalvinder and S. Grewal, “A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering,” Ver. I, 2014. [Online]. Available: www.iosrjournals.org
- [8] N. Shukla, “MEAP Edition Manning Early Access Program Machine Learning with TensorFlow,” 2017. [Online]. Available: www.manning.com
- [9] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, and C. P. Langlotz, “Deep learning in neuroradiology,” *American Journal of Neuroradiology*, vol. 39, no. 10, American Society of Neuroradiology, pp. 1776–1784, Oct. 01, 2018, doi: 10.3174/ajnr.A5543.
- [10] W. You, C. Shen, X. Guo, X. Jiang, J. Shi, and Z. Zhu, “A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery,” *Advances in Mechanical Engineering*, vol. 9, no. 6, Jun. 2017, doi: 10.1177/1687814017704146.
- [11] M. Abadi, M. Isard, and D. G. Murray, “A computational model for TensorFlow an introduction,” in MAPL 2017 - Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages, co-located with PLDI 2017, Jun. 2017, pp. 1–7, doi: 10.1145/3088525.3088527.
- [12] M. E. H. Chowdhury *et al.*, “Can AI Help in Screening Viral and COVID-19 Pneumonia?,” *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [13] T. Rahman *et al.*, “Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images,” *Computers in Biology and Medicine*, vol. 132, May 2021, doi: 10.1016/j.combiomed.2021.104319.



Penulis bernama lengkap Joshua A. Wuisan. Lahir di Manado tanggal 16 Desember 2000. Penulis tinggal di desa Sea Tumpengan, kecamatan Pineleng, Minahasa, Sulawesi Utara. Penulis menempuh pendidikan di SMA Kr. Eben Haezar Manado (2016-2018). Di tahun 2018, penulis melanjutkan Pendidikan di Universitas Sam Ratulangi dan mengambil Program Studi S-1 Teknik Informatika di Jurusan Teknik Elektro Fakultas Teknik. Selama masa perkuliahan, penulis tergabung organisasi kemahasiswaan yaitu, Unsrat IT Community (UNITY).