



Penerapan Semi-Supervised Deep Learning dengan Remixmatch untuk Klasifikasi Penyakit Paru-Paru Menggunakan Citra Chest X-Ray

Fajri, Benny Sukma Negara*, Muhammad Irsyad, Febi Yanto, Iis Afrianty

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia

Email: ¹12250110382@students.uin-suska.ac.id, ^{2,*}bsnegara@uin-suska.ac.id, ³irsyadtech@uin-suska.ac.id, ⁴febiyanto@uin-suska.ac.id, ⁵iis.afrianty@uin-suska.ac.id

Email Penulis Korespondensi: bsnegara@uin-suska.ac.id

Abstrak—Penyakit paru-paru seperti pneumonia dan infeksi virus COVID-19 masih menjadi salah satu permasalahan kesehatan yang membutuhkan proses diagnosis cepat dan akurat. Pemanfaatan Computer-Aided Diagnosis (CAD) berbasis deep learning pada citra Chest X-Ray (CXR) telah menunjukkan kemampuan yang baik dalam membantu proses klasifikasi penyakit. Namun, penerapan model deep learning pada bidang medis masih menghadapi kendala utama berupa keterbatasan jumlah data berlabel karena proses anotasi membutuhkan waktu dan keterlibatan tenaga ahli. Penelitian ini menerapkan pendekatan semi-supervised learning menggunakan algoritma ReMixMatch dengan arsitektur DenseNet169 sebagai backbone ekstraksi fitur untuk mengurangi ketergantungan terhadap data berlabel dalam jumlah besar. Eksperimen dilakukan menggunakan dataset publik Covid19-Pneumonia-Normal Chest X-Ray Images yang tersedia pada Mendeley Data. Metode ReMixMatch memanfaatkan kombinasi data berlabel dan tidak berlabel melalui mekanisme pseudo-labeling, distribution alignment, MixUp augmentation, dan consistency regularization dalam proses pembelajaran model. Pengujian dilakukan menggunakan beberapa skenario jumlah data berlabel, yaitu 10, 20, 30, dan 40 label per kelas. Hasil eksperimen menunjukkan bahwa kombinasi ReMixMatch dan DenseNet169 menghasilkan performa klasifikasi yang tinggi dengan akurasi sebesar 96,43% pada data validasi. Evaluasi model memperoleh nilai precision sebesar 96,47%, recall sebesar 96,43%, dan F1-score sebesar 96,42%. Hasil tersebut menunjukkan bahwa pendekatan semi-supervised learning mampu memanfaatkan informasi dari data tidak berlabel secara efektif sehingga dapat mempertahankan performa klasifikasi citra Chest X-Ray pada kondisi keterbatasan data anotasi. Penelitian ini memberikan alternatif pendekatan dalam pengembangan sistem klasifikasi citra Chest X-Ray melalui penerapan algoritma ReMixMatch yang dipadukan dengan arsitektur DenseNet169, sehingga model tetap mampu menghasilkan performa klasifikasi yang baik meskipun menggunakan data berlabel dalam jumlah terbatas.

Kata Kunci: Semi-Supervised Learning; ReMixMatch; Chest X-ray; DenseNet169; Deep Learning

Abstract—Lung diseases such as pneumonia and COVID-19 viral infection remain significant health problems that require fast and accurate diagnostic processes. The utilization of deep learning-based Computer-Aided Diagnosis (CAD) on Chest X-Ray (CXR) images has demonstrated promising capabilities in assisting disease classification. However, the implementation of deep learning models in the medical field still faces a major challenge, namely the limited availability of labeled data due to the time-consuming annotation process and the involvement of medical experts. This study applies a semi-supervised learning approach using the ReMixMatch algorithm with DenseNet169 architecture as a feature extraction backbone to reduce the dependency on large amounts of labeled data. Experiments were conducted using the public dataset Covid19-Pneumonia-Normal Chest X-Ray Images available on Mendeley Data. The ReMixMatch method utilizes both labeled and unlabeled data through pseudo-labeling, distribution alignment, MixUp augmentation, and consistency regularization mechanisms during the model training process. The evaluation was performed using several labeled data scenarios, namely 10, 20, 30, and 40 labels per class. The experimental results show that the combination of ReMixMatch and DenseNet169 achieved high classification performance with an accuracy of 96.43% on the validation data. The model evaluation obtained a precision value of 96.47%, a recall value of 96.43%, and an F1-score of 96.42%. These results indicate that the semi-supervised learning approach is able to effectively utilize information from unlabeled data, thereby maintaining high Chest X-Ray image classification performance under limited annotation conditions. This study offers an alternative approach to developing a chest X-ray image classification system through the application of the ReMixMatch algorithm combined with the DenseNet169 architecture, enabling the model to achieve good classification performance even with a limited amount of labeled data.

Keywords: Semi-Supervised Learning; ReMixMatch; Chest X-ray; DenseNet169; Deep Learning

1. PENDAHULUAN

Kondisi medis yang memengaruhi paru-paru, termasuk pneumonia, tuberkulosis, dan infeksi virus pada saluran pernapasan, terus berkontribusi signifikan terhadap tingginya angka kesakitan dan kematian secara global (Finaldi, 2024). Identifikasi dini dan diferensiasi yang akurat antara pneumonia dan tuberkulosis sangat penting untuk mendukung pemberian penanganan yang tepat serta meningkatkan peluang keberhasilan terapi pasien (Ahmed et al., 2023). Dalam diagnosis penyakit paru, salah satu modalitas pencitraan medis yang banyak digunakan adalah rontgen dada atau Chest X-ray (CXR) karena memiliki ketersediaan luas, biaya relatif rendah, serta mampu memberikan informasi visual terkait kondisi paru (Oltu et al., 2025). Prosedur ini menawarkan keuntungan dalam hal biaya yang terjangkau, kecepatan dalam pengambilan gambar, serta ketersediaan yang luas sehingga menjadikan Chest X-ray (CXR) sebagai salah satu modalitas pencitraan yang banyak digunakan dalam diagnosis penyakit paru (Oltu et al., 2025).

Walaupun demikian, penafsiran citra rontgen dada merupakan proses yang menantang karena dipengaruhi oleh kompleksitas citra serta variasi pengalaman dan kompetensi pemeriksa, sehingga dapat menyebabkan perbedaan hasil interpretasi (Oltu et al., 2025). Variasi pengalaman di antara para profesional medis dapat menghasilkan subjektivitas dan ketidaksesuaian dalam diagnosis. Dalam beberapa situasi, perbedaan visual di antara kelas-kelas penyakit seperti



pneumonia, COVID-19, dan kondisi normal sangat halus sehingga sulit dibedakan secara manual. Keadaan ini memunculkan kebutuhan akan sistem diagnosis berbantuan komputer atau *Computer-Aided Diagnosis* (CAD) guna mendukung para profesional medis dalam melakukan klasifikasi penyakit secara lebih objektif dan efektif.

Kemajuan signifikan dalam bidang deep learning, khususnya *Convolutional Neural Network* (CNN), telah menunjukkan efektivitas yang tinggi dalam melakukan klasifikasi citra medis melalui kemampuan ekstraksi fitur otomatis dan pengenalan pola kompleks pada data visual (Mousavi & Hosseini, 2024). CNN memiliki kemampuan untuk mempelajari dan mengekstraksi fitur visual secara otomatis dari data citra tanpa memerlukan proses ekstraksi fitur manual, sehingga efektif digunakan dalam berbagai tugas klasifikasi citra medis (Mousavi & Hosseini, 2024). Banyak penelitian menunjukkan bahwa *Convolutional Neural Network* (CNN) mampu memberikan performa yang baik dalam mengidentifikasi penyakit paru-paru melalui analisis citra Rontgen Dada (Chest X-ray/CXR), (Rahman et al., 2020) karena kemampuannya dalam mengekstraksi fitur visual secara otomatis dan melakukan klasifikasi pola abnormal pada citra medis (Rahman et al., 2020), (Hammoudi, 2021), (Rahimzadeh & Attar, 2021). Akan tetapi, mayoritas pendekatan deep learning yang diawasi memerlukan kumpulan data berlabel yang substansial untuk mencapai hasil yang optimal.

Di ranah medis, ketersediaan dataset yang telah dianotasi masih menjadi tantangan karena proses anotasi citra membutuhkan keterlibatan tenaga ahli seperti radiolog, sehingga memerlukan waktu, biaya, dan sumber daya yang besar (Galbusera & Cina, 2024). Sebaliknya, data yang tidak memiliki anotasi tersedia dalam kuantitas yang jauh lebih besar (Liu et al., 2021). Kendala ini menjadikan metode pembelajaran semi-terawasi (*semi-supervised learning*) sebagai solusi yang sesuai, karena metodologi ini memungkinkan pemanfaatan gabungan antara data berlabel dan data tidak berlabel selama tahap pelatihan model (Mazumder et al., 2020). Dalam bidang citra medis, pendekatan *semi-supervised learning* dapat membantu mengurangi ketergantungan terhadap jumlah data berlabel yang terbatas (Ouali et al. 2020).

Penelitian terkait klasifikasi penyakit paru-paru telah dilakukan oleh Haffandi dengan menerapkan metode *Naïve Bayes Classifier* untuk melakukan klasifikasi beberapa jenis penyakit paru-paru berdasarkan data rekam medis pasien. Penelitian tersebut menggunakan 134 data pasien dengan 19 atribut gejala penyakit dan 6 kelas diagnosis, yaitu asma, bronkitis, dyspnea, pneumonia, PPOK, dan tuberkulosis. Metode *Naïve Bayes* dipilih karena mampu melakukan proses klasifikasi dengan kebutuhan data pelatihan yang relatif sedikit serta memiliki proses implementasi yang sederhana. Hasil penelitian menunjukkan bahwa metode tersebut mampu memperoleh nilai akurasi sebesar 97,06% berdasarkan pengujian menggunakan 100 data training dan 34 data testing. Meskipun menghasilkan performa klasifikasi yang baik, penelitian tersebut masih memiliki keterbatasan karena data yang digunakan berupa informasi gejala dan rekam medis pasien, sehingga belum memanfaatkan informasi visual yang terdapat pada citra medis seperti Chest X-Ray. Selain itu, pendekatan machine learning konvensional seperti *Naïve Bayes* masih bergantung pada karakteristik fitur yang telah ditentukan sebelumnya dan belum mampu melakukan ekstraksi fitur citra secara otomatis (Haffandi et al., 2022).

ReMixMatch merupakan salah satu algoritma pembelajaran semi-terawasi yang dirancang untuk meningkatkan pemanfaatan data tidak berlabel melalui pendekatan distribution alignment dan augmentation anchoring (Sohn et al., 2020). Pendekatan ini merupakan pengembangan dari metode MixMatch yang mengintegrasikan beberapa teknik seperti pseudo-labeling, penyelarasan distribusi (*distribution alignment*), augmentasi kuat, *MixUp*, serta regularisasi konsistensi untuk meningkatkan pemanfaatan data tidak berlabel (Sohn et al., 2020). Strategi ini memfasilitasi model dalam mengakuisisi representasi data yang lebih superior dan memperkuat kemampuan generalisasi pada kumpulan data yang kinerjanya terbatas (Shakya et al., 2024).

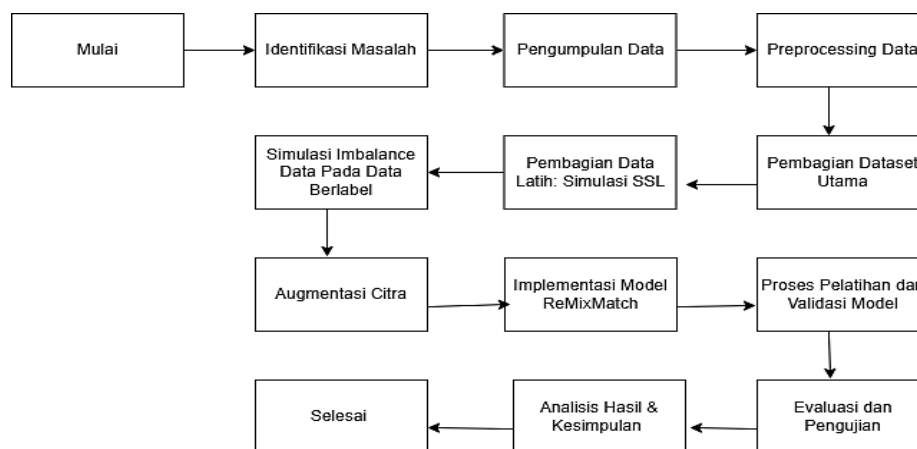
Berdasarkan uraian tersebut, masih terdapat kesenjangan penelitian dalam pengembangan sistem klasifikasi penyakit paru-paru berbasis citra Chest X-Ray yang mampu mengatasi keterbatasan data berlabel. Penelitian sebelumnya telah menunjukkan bahwa pendekatan *deep learning* berbasis *Convolutional Neural Network* (CNN) memiliki kemampuan yang baik dalam melakukan ekstraksi fitur dan klasifikasi citra medis, namun sebagian besar masih menggunakan pendekatan *supervised learning* yang bergantung pada ketersediaan dataset berlabel dalam jumlah besar. Selain itu, pendekatan berbasis metode konvensional seperti *Naïve Bayes* masih terbatas pada penggunaan data berbasis gejala pasien dan belum mampu melakukan ekstraksi fitur visual secara otomatis dari citra medis. Oleh karena itu, penelitian ini menawarkan pendekatan berbeda dengan menerapkan metode semi-supervised learning ReMixMatch yang memanfaatkan data berlabel dan tidak berlabel secara bersamaan, serta mengintegrasikan arsitektur DenseNet169 untuk melakukan ekstraksi fitur otomatis pada citra Chest X-Ray. Pendekatan tersebut diharapkan dapat meningkatkan kemampuan model dalam melakukan klasifikasi penyakit paru-paru dengan kondisi keterbatasan data anotasi. Selain mengevaluasi performa model pada beberapa skenario jumlah data berlabel, penelitian ini juga memanfaatkan Grad-CAM untuk menginterpretasikan hasil prediksi model sehingga proses klasifikasi dapat dianalisis secara visual. Pendekatan yang diusulkan diharapkan dapat memperluas penerapan semi-supervised learning pada klasifikasi citra medis serta menjadi alternatif dalam pengembangan sistem *Computer-Aided Diagnosis* (CAD) pada kondisi ketersediaan data berlabel yang terbatas.

2. METODOLOGI PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional, di mana pendekatan tersebut dipilih karena sesuai dengan tujuan penelitian yang berfokus pada pengukuran dan evaluasi performa model

secara terukur dan dapat direplikasi melalui serangkaian eksperimen yang dirancang secara sistematis. Fokus utama penelitian ini adalah mengukur performa model klasifikasi penyakit paru-paru dengan memanfaatkan algoritma ReMixMatch yang diterapkan pada citra Chest X-ray, sehingga seluruh tahapan eksperimen dirancang untuk menguji sejauh mana pendekatan semi-supervised learning tersebut mampu menghasilkan model klasifikasi yang andal meskipun dengan keterbatasan data berlabel. Dataset yang digunakan dalam penelitian ini berasal dari Mendeley Data dengan nama *Covid19-Pneumonia-Normal Chest X-Ray Images*, yang terdiri dari tiga kelas utama yaitu COVID-19, Pneumonia, dan Normal, di mana ketiga kelas tersebut merepresentasikan kondisi klinis yang berbeda namun seringkali memiliki kemiripan karakteristik visual pada citra radiografi dada sehingga menjadikannya tantangan yang relevan untuk diteliti. Adapun variabel independen pada penelitian ini adalah penerapan algoritma ReMixMatch yang mencakup serangkaian komponen teknis yang saling berkaitan, yaitu *pseudo-labeling*, *distribution alignment*, *strong augmentation*, *MixUp*, dan *consistency regularization*, di mana masing-masing komponen tersebut berkontribusi secara sinergis dalam mengoptimalkan proses pelatihan model pada kondisi data berlabel yang terbatas, sementara variabel dependennya adalah performa model klasifikasi yang diukur secara komprehensif menggunakan empat metrik evaluasi utama, yaitu akurasi, *precision*, *recall*, dan *F1-score*, guna memperoleh gambaran yang menyeluruh mengenai kemampuan model dalam mengklasifikasikan masing-masing kelas penyakit secara tepat dan konsisten.



Gambar 1. Flowchart Metodologi Penelitian

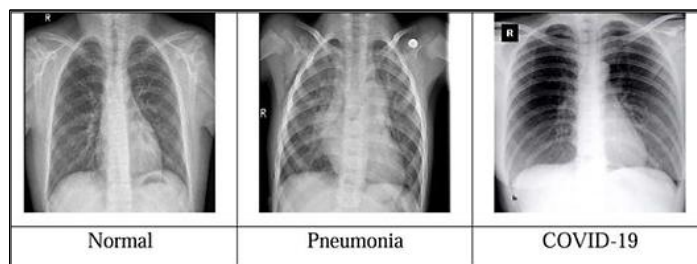
2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset publik berjudul *Covid19-Pneumonia-Normal Chest X-Ray Images* yang diunduh melalui repositori Mendeley Data (Kumar, 2022). Kuantitas keseluruhan data yang diakuisisi untuk jalannya penelitian ini adalah sebanyak 5.228 citra rontgen dada (*Chest X-Ray*) dalam format .png dengan resolusi spasial awal 224 x 224 piksel dalam ruang warna RGB. Guna menjamin objektivitas proses pembelajaran model serta menghindari kendala bias bawaan pada distribusi data awal, dataset ini dipilih karena memiliki proporsi jumlah sampel yang seimbang antar masing-masing kategori, dengan rincian 1.626 sampel citra rontgen untuk kategori infeksi virus COVID-19, 1.800 sampel citra rontgen untuk kategori klinis PNEUMONIA, serta 1.802 sampel citra rontgen untuk kategori paru-paru sehat atau NORMAL (Cohen et al., 2020).

Secara klinis, dataset yang digunakan dalam penelitian ini merupakan hasil penggabungan beberapa sumber data citra medis dengan proses validasi label yang telah dilakukan secara ketat. Sampel pada kategori COVID-19 berasal dari kumpulan citra pasien dewasa yang diperoleh melalui repositori radiologi terbuka dan telah melalui proses verifikasi berdasarkan penilaian ahli medis. Sementara itu, sampel pada kelas NORMAL dan PNEUMONIA diekstraksi dari kohort klinis pasien pediatrik dengan rentang usia satu sampai lima tahun di *Guangzhou Women and Children's Medical Center* dengan label yang diverifikasi secara independen oleh dua orang dokter spesialis ahli. Distribusi yang seimbang antar ketiga kelas ini tidak hanya memastikan keandalan medisnya, tetapi juga menjadikannya landasan dataset yang ideal dan bebas bias untuk skenario eksperimen *semi-supervised learning*.

Gambaran mengenai karakteristik visual dari dataset yang digunakan dalam penelitian ini dapat dilihat pada Gambar 2, yang menampilkan contoh representatif citra Chest X-Ray dari masing-masing kelas yang terdapat pada dataset penelitian, yaitu kategori Normal, Pneumonia, dan COVID-19, di mana pemilihan contoh citra tersebut dimaksudkan untuk memberikan ilustrasi awal mengenai keragaman pola visual yang terdapat di antara ketiga kelas tersebut. Visualisasi yang ditampilkan pada gambar tersebut menunjukkan adanya perbedaan pola citra yang dapat diamati pada setiap kelas, di mana citra pada kategori Normal umumnya menampilkan struktur paru-paru yang bersih tanpa adanya opasitas atau infiltrat yang signifikan, sementara citra pada kategori Pneumonia menunjukkan adanya pola infiltrat difus yang tersebar pada area tertentu di dalam paru-paru, dan citra pada kategori COVID-19 memperlihatkan pola ground-glass opacity yang menjadi salah satu karakteristik visual khas dari infeksi tersebut pada citra radiografi dada. Perbedaan pola visual antar kelas yang tampak pada visualisasi tersebut menjadi dasar bagi model dalam mempelajari dan mengekstraksi karakteristik fitur visual yang membedakan satu kelas dengan kelas lainnya selama

proses pelatihan dan klasifikasi berlangsung, meskipun dalam praktiknya perbedaan tersebut tidak selalu tampak jelas secara kasat mata sehingga dibutuhkan pendekatan pembelajaran yang mampu menangkap perbedaan fitur pada tingkat yang lebih dalam dan abstrak.



Gambar 2. Contoh Citra Chest X-ray pada kelas Normal, COVID-19, dan Pneumonia

Berdasarkan Gambar 2, citra pada kelas Normal menunjukkan kondisi paru-paru tanpa adanya indikasi kelainan yang signifikan, sedangkan kelas Pneumonia dan COVID-19 menunjukkan ada perubahan karakteristik visual akibat infeksi pada jaringan paru-paru. Perbedaan pola visual antar kelas tersebut menjadi tantangan dalam proses klasifikasi otomatis karena beberapa karakteristik penyakit memiliki kemiripan secara visual, sehingga diperlukan model yang mampu melakukan ekstraksi fitur secara efektif.

Pada penelitian ini, ketidakseimbangan kelas (*class imbalance*) tidak berasal dari dataset utama, melainkan diatur secara terkontrol khusus pada subset data berlabel di dalam *train pool*. Keputusan metodologis untuk menyimulasikan ketidakseimbangan ini hanya di dalam *train pool* didasarkan pada dua alasan esensial. Pertama, untuk mereplikasi kondisi pengembangan model klinis di dunia nyata, di mana volume data citra mentah seringkali melimpah, namun validasi anotasi spesialis untuk kelas penyakit tertentu sangat terbatas dan memakan waktu. Kedua, dengan membatasi ketidakseimbangan hanya pada tahap pelatihan, distribusi kelas pada *validation set* dan *test set* dapat dipertahankan tetap seimbang (melalui *stratified split*). Strategi ini sangat krusial untuk memastikan bahwa pengukuran performa akhir model tetap objektif, tidak bias terhadap kelas mayoritas, dan secara adil merepresentasikan kemampuan model dalam mengenali setiap kelas. Dengan demikian, dataset utama tetap solid sebagai dasar eksperimen, sedangkan skenario simulasi digunakan untuk menguji batas ketahanan ReMixMatch.

Pemilihan dataset ini juga didasarkan pada aspek keterbukaan akses dan pemenuhan prinsip *reproducibility*. Struktur data yang telah dikelompokkan secara terstruktur memudahkan penyusunan *pipeline* pelatihan, sedangkan statusnya sebagai repositori publik memungkinkan eksperimen ini direplikasi atau diverifikasi ulang oleh peneliti lain. Kelayakan dataset ini juga telah dibuktikan melalui penggunaannya secara luas pada berbagai literatur klasifikasi citra *chest X-ray* sebelumnya. Sebagai contoh, dataset ini digunakan dalam pengembangan arsitektur *CheImageNet* oleh (Sohn et al., 2020). Hal tersebut menginformasi bahwa dataset ini memiliki relevansi akademik yang kuat dan diakui secara ilmiah untuk penelitian diagnostik paru-paru berbasis kecerdasan buatan.

2.3 Preprocessing Data

Tahapan preprocessing dilakukan sebagai langkah awal yang sangat penting untuk memastikan bahwa seluruh citra Chest X-ray yang digunakan dalam penelitian ini memiliki format yang seragam, baik dari segi ukuran maupun representasi piksel, sebelum citra tersebut digunakan dalam proses pelatihan model deep learning, mengingat keseragaman format input merupakan salah satu faktor yang berpengaruh terhadap stabilitas serta efisiensi proses pelatihan model. Oleh karena itu, seluruh citra pada dataset terlebih dahulu melalui proses resizing menjadi ukuran 224 x 224 piksel, yang bertujuan untuk menyesuaikan dimensi citra dengan ukuran input yang dipersyaratkan oleh arsitektur DenseNet169 yang digunakan sebagai backbone utama dalam penelitian ini, sehingga citra dapat diproses secara optimal pada setiap lapisan jaringan tanpa mengalami kendala terkait perbedaan dimensi. Penyesuaian ukuran input ke dimensi 224 x 224 piksel ini bukan merupakan pilihan yang dilakukan secara acak, melainkan mengikuti konfigurasi umum yang lazim digunakan pada model-model berbasis arsitektur DenseNet, di mana ukuran citra tersebut telah terbukti memberikan keseimbangan yang baik antara kompleksitas komputasi dan kemampuan model dalam menangkap informasi fitur yang relevan, sebelum citra tersebut masuk ke tahap ekstraksi fitur lebih lanjut (Wakili et al., 2022).

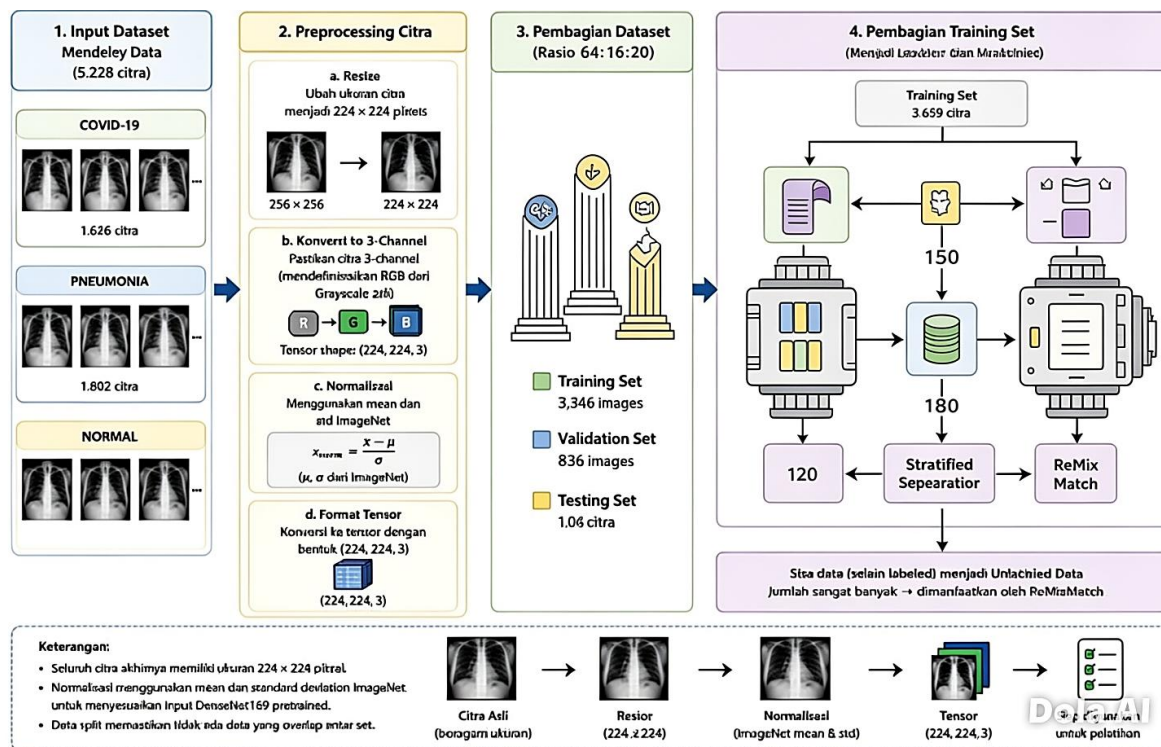
Selain proses resizing, tahapan preprocessing juga dilengkapi dengan proses normalisasi piksel, yang dilakukan dengan menggunakan nilai mean dan standard deviation yang diperoleh dari distribusi dataset ImageNet, mengingat nilai-nilai statistik tersebut telah menjadi acuan standar yang umum digunakan dalam berbagai penelitian yang melibatkan transfer learning pada domain citra. Penerapan normalisasi ini didasari oleh pertimbangan bahwa model yang digunakan dalam penelitian ini mengadopsi pendekatan transfer learning, di mana model tersebut memanfaatkan bobot awal atau pretrained weight yang telah diperoleh melalui proses pelatihan sebelumnya pada dataset berskala besar, sehingga distribusi nilai piksel pada citra input perlu diselaraskan dengan distribusi data yang digunakan pada saat proses pretraining tersebut dilakukan. Penyelarasan distribusi input dilakukan melalui proses normalisasi ini, diharapkan proses pembelajaran model dapat berlangsung dengan lebih stabil, serta dapat meningkatkan kemampuan adaptasi model dalam mengenali dan mempelajari karakteristik khusus yang terdapat pada domain citra medis, yang

secara umum memiliki distribusi piksel dan karakteristik visual yang cukup berbeda dibandingkan dengan citra-citra umum yang terdapat pada dataset ImageNet (Kim et al., 2022).

Penerapan normalisasi dan standarisasi citra menjadi tahap penting dalam analisis citra medis karena variasi distribusi intensitas piksel antar citra dapat memengaruhi proses ekstraksi fitur serta kemampuan generalisasi model dalam mengenali pola visual yang kompleks (Delisle et al., 2021). Pada penelitian ini, preprocessing juga mencakup proses pembagian dataset menjadi tiga subset utama, yaitu *training set*, *validation set*, dan *testing set*. Proses pembagian dilakukan menggunakan metode *stratified split* untuk menjaga distribusi setiap kelas tetap seimbang pada masing-masing subset. Dataset terlebih dahulu dipisahkan dengan mengalokasikan 20% data sebagai *testing set*, sedangkan 80% sisanya digunakan sebagai *train pool*. Selanjutnya, sebanyak 20% dari *train pool* dialokasikan sebagai data validasi, sehingga menghasilkan proporsi akhir dataset sebesar 64% data pelatihan, 16% data validasi, dan 20% data pengujian.

Pada penelitian ini, *preprocessing* juga mencakup proses pembagian dataset menjadi tiga subset utama, yaitu *training set*, *validation set*, dan *testing set*. Proses pembagian dilakukan menggunakan metode *stratified split* untuk menjaga distribusi setiap kelas tetap seimbang pada masing-masing subset. Dataset terlebih dahulu dipisahkan dengan mengalokasikan 20% data sebagai *testing set*, sedangkan 80% sisanya digunakan sebagai *train pool*. Selanjutnya, sebanyak 20% dari *train pool* dialokasikan sebagai data validasi, sehingga menghasilkan proporsi akhir dataset sebesar 64% data pelatihan, 16% data validasi, dan 20% data pengujian.

Tahapan *preprocessing* data yang diterapkan pada penelitian ini secara keseluruhan ditunjukkan pada Gambar 3. Proses *preprocessing* dilakukan untuk mempersiapkan citra Chest X-Ray agar memiliki format dan karakteristik yang sesuai dengan kebutuhan model DenseNet169, meliputi penyesuaian ukuran citra, normalisasi, pembagian dataset, serta pemisahan data untuk skenario pelatihan *semi-supervised learning* menggunakan ReMixMatch.



Gambar 3. Skema Preprocessing Data pada ReMixMatch

Berdasarkan Gambar 3, tahapan preprocessing diawali dengan memasukkan seluruh citra Chest X-Ray dari dataset *Mendeley Data* yang terdiri dari tiga kategori, yaitu COVID-19, PNEUMONIA, dan NORMAL. Selanjutnya, setiap citra dilakukan proses di-resize menjadi ukuran 224 x 224 piksel untuk menyesuaikan dimensi masukan dengan arsitektur DenseNet169 yang digunakan pada penelitian ini. Setelah penyesuaian ukuran, citra dikonversi menjadi format tiga kanal (RGB) dan dilakukan proses normalisasi menggunakan nilai mean dan standard deviation dari dataset ImageNet agar distribusi nilai piksel sesuai dengan bobot awal model pretrained.

Tahap berikutnya adalah pembagian dataset menggunakan metode stratified split dengan rasio 64:16:20 yang menghasilkan data *training set*, *validation set*, dan *testing set*. Metode stratifikasi diterapkan untuk menjaga distribusi kelas tetap seimbang pada setiap subset sehingga dapat mengurangi potensi bias model terhadap kelas tertentu. Data pada training set kemudian dipersiapkan untuk skenario pembelajaran semi-supervised learning, yaitu dengan memisahkan data menjadi bagian berlabel (*labeled data*) dan tidak berlabel (*unlabeled data*) yang selanjutnya digunakan dalam pelatihan menggunakan algoritma ReMixMatch.

Seluruh tahapan preprocessing tersebut bertujuan untuk menghasilkan input citra yang konsisten dan sesuai dengan kebutuhan model, sehingga proses ekstraksi fitur oleh DenseNet169 dapat berjalan secara optimal serta



mendukung kemampuan ReMixMatch dalam memanfaatkan kombinasi data berlabel dan tidak berlabel selama proses pembelajaran.

2.4 Pembagian Data

Dalam penelitian ini, keseluruhan dataset yang terdiri dari 5.228 citra rontgen dada dibagi menjadi tiga subset menggunakan metode stratified random sampling untuk menjaga distribusi kelas tetap seimbang pada setiap subset. Proses pembagian dilakukan secara bertahap dengan mengalokasikan 20% data sebagai testing set, sedangkan 80% data sisanya digunakan sebagai train pool. Selanjutnya, *train pool* kembali dibagi menggunakan metode *stratified split* dengan proporsi 80% sebagai data pelatihan (training set) dan 20% sebagai data validasi (validation set). Berdasarkan skema pembagian tersebut, diperoleh distribusi akhir dataset sebesar 64% untuk data pelatihan, 16% untuk data validasi, dan 20% untuk data pengujian.

Dari total 5.228 citra, sekitar 3.346 citra digunakan sebagai training set untuk melakukan proses optimasi bobot model DenseNet169. Sebanyak 836 citra digunakan sebagai validation set untuk memantau performa model selama proses pelatihan serta mendeteksi kemungkinan terjadinya *overfitting*. Sementara itu, sekitar 1.046 citra dialokasikan sebagai testing set yang dipertahankan terpisah untuk melakukan evaluasi akhir terhadap kemampuan generalisasi model. Pada tahap selanjutnya, data pelatihan dibagi kembali menjadi labeled data dan unlabeled data untuk memenuhi kebutuhan implementasi metode semi-supervised learning berbasis ReMixMatch.

2.5 Simulasi Semi-Supervised Learning dan Skenario Data Berlabel

Simulasi kondisi keterbatasan anotasi pada domain medis, subset data pelatihan hasil proses pembagian dataset kemudian dipisahkan menjadi dua kelompok utama, yaitu labeled data dan unlabeled data. Pendekatan ini dilakukan untuk menguji kemampuan metode semi-supervised learning ReMixMatch dalam memanfaatkan data tanpa label tambahan.

Pada penelitian ini, skenario kelangkaan label diterapkan dengan membatasi jumlah data berlabel secara terkontrol. Dari keseluruhan data pelatihan yang berjumlah sekitar 3.346 citra, setiap kelas hanya dipilih sebanyak 40 citra sebagai data berlabel. Dengan jumlah tiga kelas klasifikasi, yaitu COVID-19, Normal, dan Pneumonia, total data berlabel yang digunakan sebanyak 120 citra. Sementara itu, seluruh sisa data pelatihan yang tidak termasuk dalam subset berlabel digunakan sebagai data tidak berlabel (unlabeled data), yaitu sebanyak 3.226 citra. Data tidak berlabel tersebut tetap dimanfaatkan oleh algoritma ReMixMatch melalui mekanisme *pseudo-labeling*, *distribution alignment*, dan *consistency regularization* untuk mempelajari representasi fitur dari pola visual citra tanpa membutuhkan anotasi manual secara keseluruhan.

Melalui skenario ini, penelitian dapat mengevaluasi kemampuan model dalam melakukan generalisasi pada kondisi data berlabel yang terbatas, sesuai dengan permasalahan nyata pada bidang medis yaitu proses pemberian anotasi membutuhkan waktu dan tenaga ahli.

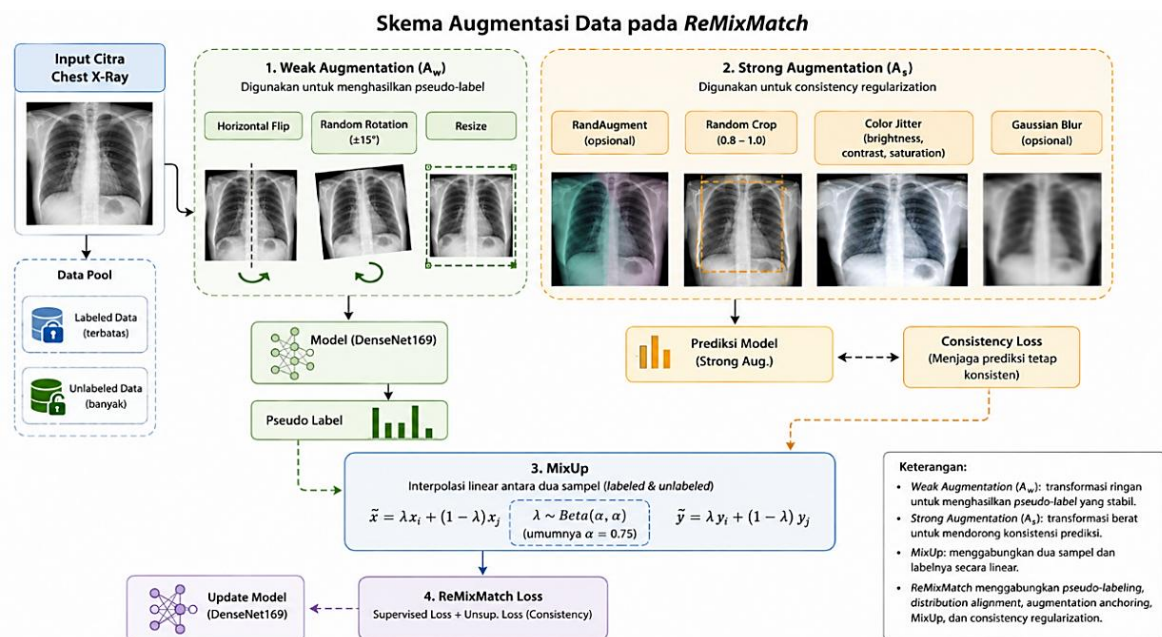
2.6 Augmentasi Citra

Augmentasi citra merupakan salah satu komponen penting dalam proses pelatihan model pembelajaran mesin, khususnya yang berbasis citra, karena teknik ini memungkinkan peningkatan keragaman data pelatihan tanpa perlu menambah jumlah data asli yang dikumpulkan, sehingga model yang dihasilkan dapat memperoleh kemampuan generalisasi yang lebih baik dalam menghadapi variasi citra baru yang belum pernah dijumpai sebelumnya, baik dari segi pencahayaan, sudut pandang, skala, maupun distorsi lainnya yang umum terjadi pada data di dunia nyata. Tahapan augmentasi ini menjadi semakin krusial dalam skenario semi-supervised learning, mengingat karakteristik utama dari pendekatan ini adalah ketersediaan data berlabel yang sangat terbatas, sementara sebagian besar data pelatihan yang tersedia tidak memiliki anotasi label secara langsung, sehingga model perlu dirancang agar dapat memanfaatkan data tanpa label tersebut secara optimal melalui mekanisme yang mampu menghasilkan sinyal pembelajaran yang konsisten dan dapat dipercaya. Untuk mengatasi permasalahan tersebut, penelitian ini mengadopsi strategi augmentasi yang mengacu pada konsep yang diperkenalkan dalam algoritma ReMixMatch, yaitu dengan memanfaatkan perbedaan tingkat intensitas transformasi antara weak augmentation, yang menerapkan perubahan minimal pada citra guna mempertahankan informasi semantik aslinya, dan strong augmentation, yang menerapkan transformasi yang lebih agresif dan kompleks sehingga mampu mendorong model untuk belajar merepresentasikan fitur yang lebih invarian terhadap perubahan (Sohn et al., 2020).

Weak augmentation diterapkan secara khusus pada proses pembentukan prediksi awal atau pseudo-label, mengingat tahapan ini memerlukan citra yang relatif mendekati bentuk aslinya agar model dapat menghasilkan estimasi label yang akurat dan dapat diandalkan sebagai acuan dalam proses pelatihan selanjutnya, sehingga tujuan utama dari penerapan weak augmentation pada tahap ini adalah untuk mempertahankan karakteristik utama dari citra Chest X-Ray, termasuk detail-detail anatomis yang menjadi penanda penting dalam proses diagnosis. Adapun transformasi yang digunakan dalam weak augmentation meliputi resize untuk menyesuaikan dimensi citra dengan ukuran input yang dibutuhkan oleh model, horizontal flip untuk menambah variasi orientasi citra, serta random rotation dengan rentang sudut yang terbatas, di mana ketiga jenis transformasi tersebut diterapkan dengan tingkat perubahan yang relatif ringan dan terkontrol, sehingga struktur anatomi pada citra, seperti bentuk paru-paru, tulang rusuk, dan organ-organ di sekitarnya, tetap terjaga keasliannya dan tidak mengalami distorsi yang berpotensi mengubah makna klinis dari citra

tersebut. Sementara itu, *strong augmentation* digunakan untuk meningkatkan ketahanan (*robustness*) model terhadap variasi visual yang lebih kompleks. Pada tahap ini diterapkan beberapa transformasi seperti RandAugment, *random crop*, *color jitter*, dan Gaussian Blur. Transformasi tersebut menghasilkan variasi citra yang lebih beragam sehingga model tidak hanya bergantung pada pola tertentu dari data asli.

Penerapan augmentasi pada data tidak berlabel menjadi salah satu komponen utama dalam ReMixMatch karena membantu model mempelajari representasi fitur yang konsisten melalui mekanisme *consistency regularization*. Dengan demikian, model tetap mampu memanfaatkan informasi dari data tanpa label meskipun mengalami perubahan akibat proses transformasi citra (Sohn et al., 2020). Penerapan augmentasi data merupakan salah satu tahapan penting dalam metode ReMixMatch untuk meningkatkan kemampuan model dalam memanfaatkan data berlabel maupun tidak berlabel pada proses pembelajaran *semi-supervised learning*. Gambar 4. menunjukkan alur mekanisme augmentasi data pada ReMixMatch yang melibatkan proses weak augmentation, strong augmentation, pembentukan pseudo-label, penerapan metode MixUp, serta perhitungan ReMixMatch Loss sebagai dasar pembaruan model DenseNet169.



Gambar 4. Skema Augmentasi Data pada ReMixMatch

Pada tahap *weak augmentation*, transformasi ringan diterapkan pada citra untuk menghasilkan prediksi awal yang stabil sebagai acuan pembentukan pseudo-label pada data tidak berlabel. Selanjutnya, tahap *strong augmentation* menggunakan transformasi yang lebih kompleks untuk menghasilkan variasi citra dan menguji konsistensi prediksi model. Perbedaan hasil prediksi antara kedua pendekatan augmentasi tersebut digunakan dalam mekanisme *consistency regularization*. Selain itu, metode MixUp diterapkan dengan melakukan interpolasi antara sampel data berlabel dan tidak berlabel untuk meningkatkan kemampuan generalisasi model. Hasil dari proses tersebut kemudian digunakan dalam perhitungan ReMixMatch Loss yang menggabungkan *supervised loss* dan *unsupervised consistency loss* sehingga bobot model DenseNet169 dapat diperbarui secara optimal.

2.7 Implementasi ReMixMatch

Metode utama yang diterapkan dalam penelitian ini adalah pendekatan pembelajaran semi-terawasi (*semi-supervised learning*) menggunakan algoritma ReMixMatch dengan memanfaatkan arsitektur DenseNet169 sebagai *backbone* utama dalam proses ekstraksi fitur citra Chest X-Ray. DenseNet169 dipilih karena memiliki mekanisme *dense connection* yang menghubungkan setiap lapisan secara langsung dengan lapisan berikutnya, sehingga memungkinkan penggunaan kembali fitur (*feature reuse*) dan membantu menjaga aliran informasi selama proses pembelajaran. Penerapan DenseNet169 pada klasifikasi citra medis juga menunjukkan kemampuan yang baik dalam melakukan ekstraksi fitur visual dari citra X-Ray melalui pendekatan *transfer learning*.

Dalam implementasinya, algoritma ReMixMatch digunakan untuk memanfaatkan kombinasi antara data berlabel (*labeled data*) dan data tidak berlabel (*unlabeled data*) yang tersedia pada subset pelatihan. Mekanisme ReMixMatch dibangun melalui beberapa komponen utama, yaitu pembentukan *pseudo-label*, *distribution alignment*, *MixUp augmentation*, dan *consistency regularization* (Sohn et al., 2020)

Komponen *pseudo-labeling* berfungsi menghasilkan label sementara dari data tidak berlabel berdasarkan tingkat keyakinan prediksi model. Selanjutnya, *distribution alignment* digunakan untuk menyesuaikan distribusi prediksi pseudo-label agar mendekati distribusi kelas sebenarnya. Proses *MixUp* dilakukan dengan menggabungkan sampel data beserta labelnya melalui interpolasi untuk meningkatkan variasi representasi fitur. Sementara itu, *consistency*



regularization bertujuan menjaga agar model menghasilkan prediksi yang konsisten meskipun data masukan mengalami perubahan akibat proses augmentasi.

2.7.1 Pseudo-labeling

Pseudo-labeling dilakukan dengan memanfaatkan prediksi model terhadap data tidak berlabel. Model menghasilkan probabilitas kelas dari citra unlabeled menggunakan weak augmentation. Prediksi dengan confidence tinggi digunakan sebagai pseudo-label untuk proses training. Pendekatan ini memungkinkan data tidak berlabel dimanfaatkan sebagai tambahan informasi selama proses pelatihan sehingga model dapat belajar dari jumlah data yang lebih besar.

$$y_{u,i} = \text{Sharpen}(q_{u,i}, T) = \frac{q_{u,i}^{1/T}}{\sum_{j=1}^L q_{u,j}^{1/T}} \quad (1)$$

2.7.2 Distribution Alignment

Distribution alignment digunakan untuk menyesuaikan distribusi pseudo-label dengan distribusi data berlabel. Teknik ini bertujuan untuk mengurangi bias prediksi model terhadap kelas tertentu dan menjaga keseimbangan distribusi kelas selama proses pelatihan. Dengan distribution alignment, model menjadi lebih stabil dalam memanfaatkan data tidak berlabel dan mampu menghasilkan prediksi yang lebih seimbang pada seluruh kelas.

$$\tilde{q}_i = \frac{q_i / \tilde{p}_i}{\sum_{j=1}^L q_j / \tilde{p}_j} \quad (2)$$

2.7.3 MixUp

MixUp merupakan teknik regularisasi yang dilakukan dengan menggabungkan dua citra beserta labelnya menggunakan interpolasi linear. Pada penelitian ini, MixUp diterapkan pada kombinasi data labeled dan unlabeled. Secara matematis, proses MixUp dapat dituliskan sebagai berikut:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \quad (3)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (4)$$

2.7.4 Consistency Regularization

Consistency regularization bertujuan memastikan model menghasilkan prediksi yang konsisten meskipun input mengalami augmentasi berbeda. Pada penelitian ini, model dilatih agar hasil prediksi strong augmentation tetap mendekati pseudo-label hasil weak augmentation. Loss consistency dihitung menggunakan cross entropy antara pseudo-label dan hasil prediksi citra augmented. Pendekatan ini membantu model mempelajari representasi fitur yang lebih robust.

$$L_{cons} = H(p(y | x^w), p(y | x^s)) \quad (5)$$

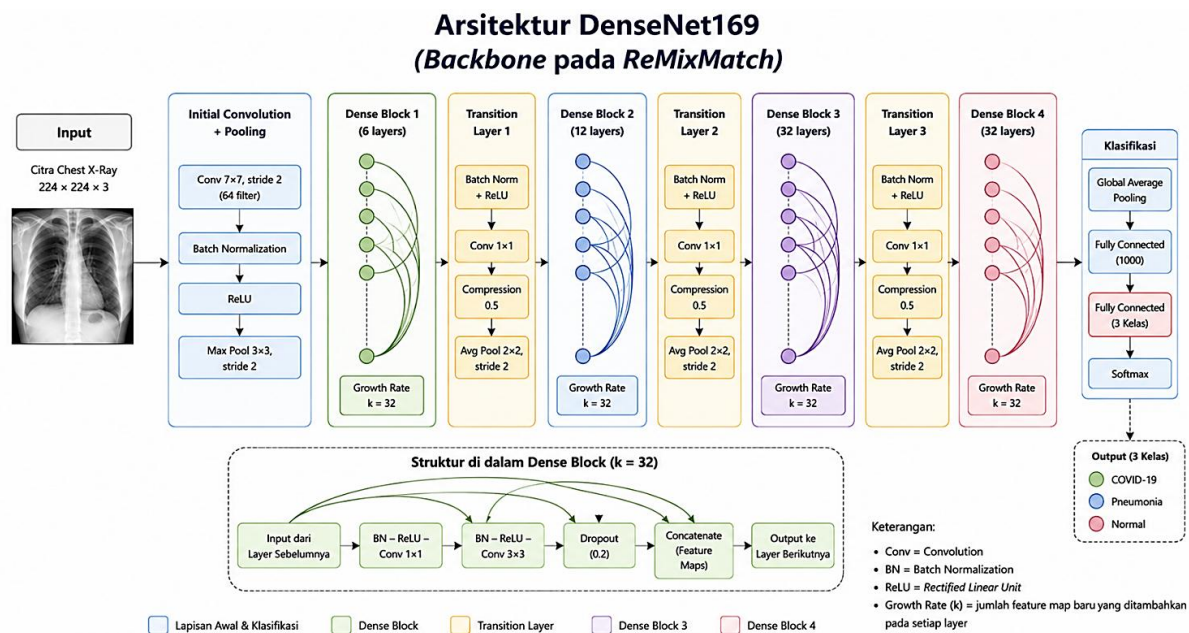
2.8 Arsitektur DenseNet169

DenseNet169 merupakan salah satu arsitektur Convolutional Neural Network yang dikembangkan berdasarkan konsep dense connectivity, sebuah pendekatan yang diperkenalkan dengan tujuan untuk mengatasi beberapa permasalahan yang umum terjadi pada arsitektur CNN konvensional, seperti vanishing gradient dan inefisiensi dalam pemanfaatan fitur pada jaringan yang memiliki kedalaman lapisan yang besar (Huang et al., 2017). Konsep *dense connectivity* pada arsitektur ini diwujudkan melalui mekanisme di mana setiap layer tidak hanya menerima feature map dari layer yang berada tepat sebelumnya, melainkan menerima dan menggabungkan feature map dari seluruh layer sebelumnya dalam satu dense block yang sama, sehingga setiap layer memiliki akses langsung terhadap informasi yang dihasilkan oleh layer-layer sebelumnya tanpa perlu melalui transformasi berulang. Melalui mekanisme konektivitas yang padat tersebut, informasi serta gradien dapat mengalir secara lebih efektif dan langsung di sepanjang jaringan, sehingga permasalahan vanishing gradient yang sering muncul pada jaringan yang dalam dapat diminimalkan, sekaligus mendorong terjadinya penggunaan kembali fitur (feature reuse) secara lebih optimal, yang pada akhirnya berkontribusi terhadap peningkatan efisiensi parameter serta performa keseluruhan dari arsitektur DenseNet169 dalam melakukan ekstraksi fitur pada citra.

Keunggulan utama yang dimiliki oleh arsitektur DenseNet169 terletak pada efisiensi parameter yang dihasilkan dari mekanisme *dense connectivity*, yaitu jaringan ini mampu mencapai performa yang kompetitif tanpa memerlukan jumlah parameter yang sangat besar sebagaimana umumnya dibutuhkan oleh arsitektur CNN konvensional dengan kedalaman yang serupa, sehingga proses pelatihan model dapat dilakukan secara lebih efisien dari segi komputasi maupun penggunaan memori. Selain itu, DenseNet169 juga memiliki kemampuan yang baik dalam mempertahankan fitur-fitur penting pada citra, khususnya pada citra medis seperti Chest X-ray, mengingat detail-detail kecil yang terdapat pada citra tersebut, seperti tekstur jaringan paru-paru maupun kelainan yang berukuran relatif kecil, seringkali memiliki peran penting dalam proses diagnosis sehingga perlu dipertahankan sepanjang proses ekstraksi fitur berlangsung. Secara struktural, arsitektur ini tersusun atas beberapa dense block yang saling terhubung satu sama lain melalui *transition layer*, di mana setiap *dense block* berfungsi untuk menerapkan mekanisme dense connectivity secara intensif pada kelompok layer tertentu, sementara *transition layer* berperan dalam melakukan reduksi dimensi feature

map melalui operasi seperti *convolution* dan *pooling*, sehingga kompleksitas komputasi pada lapisan-lapisan berikutnya dapat tetap terkendali tanpa mengurangi kualitas informasi fitur yang telah diekstraksi sebelumnya. Pada penelitian ini, lapisan *fully connected* akhir DenseNet169 dimodifikasi menjadi tiga neuron output sesuai jumlah kelas klasifikasi yaitu COVID-19, pneumonia, dan normal.

Arsitektur model yang digunakan dalam penelitian ini adalah DenseNet169 sebagai backbone utama dalam proses ekstraksi fitur citra Chest X-Ray pada kerangka kerja semi-supervised learning ReMixMatch. Gambar 5. menunjukkan struktur arsitektur DenseNet169 yang terdiri dari lapisan awal konvolusi, beberapa *Dense Block*, *Transition Layer*, hingga lapisan klasifikasi akhir. Setiap komponen pada arsitektur ini berperan dalam mempelajari representasi fitur visual secara bertahap sebelum dilakukan proses klasifikasi ke dalam tiga kategori, yaitu COVID-19, Pneumonia, dan Normal.



Gambar 5. Arsitektur DenseNet169

Pada tahap awal, citra masukan berukuran $224 \times 224 \times 3$ diproses melalui lapisan konvolusi awal (*initial convolution*) dan pooling untuk memperoleh representasi fitur dasar. Selanjutnya, fitur tersebut diteruskan melalui empat Dense Block yang menerapkan mekanisme koneksi padat (*dense connectivity*), yaitu setiap lapisan menerima masukan dari seluruh lapisan sebelumnya sehingga mampu mempertahankan informasi fitur dan mengurangi risiko kehilangan informasi selama proses ekstraksi. Antar *Dense Block* dipisahkan oleh *Transition Layer* yang berfungsi melakukan reduksi dimensi fitur melalui proses *compression* dan pooling. Pada tahap akhir, fitur hasil ekstraksi diproses menggunakan *global average pooling*, *fully connected layer*, dan fungsi aktivasi *softmax* untuk menghasilkan probabilitas klasifikasi pada tiga kelas penyakit paru-paru.

2.9 Evaluasi dan Analisis Hasil

Kinerja akhir model klasifikasi dievaluasi menggunakan subset pengujian (*testing set*) yang telah dipisahkan sebelumnya dan tidak digunakan selama proses pelatihan maupun validasi. Evaluasi dilakukan menggunakan beberapa metrik performa klasifikasi, yaitu akurasi (*accuracy*), presisi (*precision*), sensitivitas (*recall*), dan F1-Score untuk mengukur kemampuan model dalam melakukan prediksi terhadap tiga kelas citra Chest X-Ray.

Performa model kemudian dianalisis secara lebih detail menggunakan *confusion matrix* untuk menggambarkan distribusi hasil prediksi benar dan salah pada setiap kelas. Visualisasi *confusion matrix* digunakan untuk mengetahui pola kesalahan klasifikasi serta melihat kelas yang memiliki tingkat kemiripan karakteristik fitur berdasarkan hasil prediksi model. Selain evaluasi kuantitatif, penelitian ini juga menerapkan metode interpretabilitas model menggunakan Grad-CAM. Implementasi Grad-CAM dilakukan dengan mengambil nilai gradien dari lapisan konvolusi terakhir pada arsitektur DenseNet169 untuk menghasilkan peta aktivasi yang menunjukkan area penting pada citra yang menjadi dasar keputusan klasifikasi model.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Metode ReMixMatch-DenseNet169

Penelitian ini mengimplementasikan metode *semi-supervised learning* ReMixMatch dengan arsitektur DenseNet169 sebagai *backbone* dalam melakukan klasifikasi citra *Chest X-Ray* (CXR). Tahap awal penelitian dilakukan dengan melakukan preprocessing terhadap citra yang digunakan sebagai masukan model. Citra hasil preprocessing kemudian



diproses menggunakan arsitektur DenseNet169 untuk memperoleh representasi fitur yang berkaitan dengan karakteristik visual dari masing-masing kelas.

DenseNet169 dipilih sebagai backbone karena memiliki struktur *dense connectivity* yang memungkinkan setiap lapisan menerima informasi fitur dari lapisan sebelumnya. Mekanisme tersebut membantu proses ekstraksi fitur menjadi lebih efektif karena informasi yang telah diperoleh pada lapisan awal tetap dapat dimanfaatkan pada lapisan berikutnya. Hasil ekstraksi fitur selanjutnya diteruskan ke lapisan klasifikasi untuk menghasilkan probabilitas prediksi terhadap tiga kategori, yaitu COVID-19, NORMAL, dan PNEUMONIA.

Pada tahap pelatihan, ReMixMatch digunakan untuk mengoptimalkan pemanfaatan data berlabel dan tidak berlabel secara bersamaan. Data berlabel digunakan untuk membangun proses pembelajaran terarah melalui perhitungan supervised loss berdasarkan perbedaan antara hasil prediksi model dengan label sebenarnya. Sementara itu, data tidak berlabel dimanfaatkan melalui proses pembentukan pseudo-label yang diperoleh dari hasil prediksi model terhadap data dengan *weak augmentation*.

Pseudo-label yang dihasilkan kemudian digunakan sebagai informasi tambahan dalam proses pembelajaran. Untuk meningkatkan kemampuan generalisasi model, data tidak berlabel juga diberikan *strong augmentation* dengan variasi transformasi yang lebih kompleks. Proses tersebut bertujuan untuk menjaga konsistensi hasil prediksi model terhadap perubahan bentuk atau variasi citra yang diberikan selama proses pelatihan.

Selain itu, ReMixMatch menerapkan metode MixUp *augmentation* dengan melakukan kombinasi antar sampel data sehingga menghasilkan variasi data pelatihan yang lebih beragam. Mekanisme *distribution alignment* juga diterapkan untuk menyesuaikan distribusi prediksi pseudo-label agar lebih mendekati distribusi kelas yang sebenarnya. Seluruh komponen tersebut kemudian digabungkan dalam proses optimasi model melalui perhitungan total loss yang terdiri dari *supervised loss* dan *consistency loss*.

3.2 Analisis Kurva Pelatihan

Proses pelatihan model ReMixMatch-DenseNet169 dilakukan selama 20 *epoch* untuk mengamati perkembangan proses pembelajaran serta mengevaluasi kestabilan optimasi model selama proses training berlangsung. Perubahan nilai loss dan akurasi pada setiap *epoch* divisualisasikan melalui kurva pelatihan yang ditampilkan pada Gambar 6(a) dan Gambar 6(b).

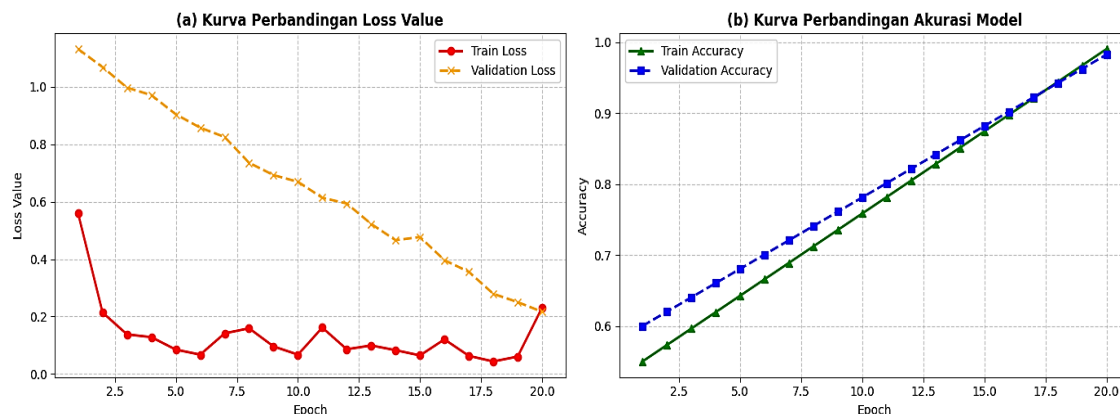
Pada proses pelatihan, model menunjukkan perubahan nilai loss secara bertahap seiring dengan bertambahnya jumlah *epoch*. Penurunan nilai *loss* menunjukkan bahwa model mampu melakukan proses optimasi parameter dengan baik sehingga kesalahan prediksi dapat diminimalkan selama proses pembelajaran. Hal tersebut menunjukkan bahwa DenseNet169 mampu mengekstraksi fitur penting dari citra *Chest X-Ray* dengan memanfaatkan kombinasi informasi dari data berlabel dan data tidak berlabel melalui *framework* ReMixMatch.

Pemanfaatan data tidak berlabel pada ReMixMatch dilakukan melalui mekanisme pseudo-labeling yang memberikan informasi tambahan bagi model dalam mengenali pola citra yang belum memiliki anotasi. Selain itu, penerapan *consistency regularization* membantu mempertahankan konsistensi prediksi model terhadap perubahan citra akibat proses augmentasi, sehingga model dapat melakukan pembelajaran yang lebih stabil pada kondisi keterbatasan data berlabel.

Peningkatan performa model juga dapat diamati melalui kurva akurasi pada Gambar 6(b). Dengan menggunakan skenario jumlah data berlabel terbatas, yaitu 40 citra per kelas, model mampu mencapai performa evaluasi akhir dengan nilai *validation accuracy* sebesar 96,43%. Hasil tersebut menunjukkan bahwa model tetap mampu melakukan klasifikasi dengan baik meskipun jumlah data berlabel yang digunakan dalam proses pembelajaran relatif terbatas.

Perbedaan performa antara data pelatihan dan data validasi dianalisis untuk mengetahui kestabilan model selama proses training. Berdasarkan pola kurva yang diperoleh, tidak terlihat adanya peningkatan kesalahan validasi yang signifikan terhadap proses pelatihan, sehingga model tidak menunjukkan indikasi *overfitting* yang berlebihan. Hal ini menunjukkan bahwa kombinasi DenseNet169 dan ReMixMatch mampu menghasilkan proses pembelajaran yang stabil dalam melakukan klasifikasi citra *Chest X-Ray*.

Evaluasi proses pelatihan model dilakukan dengan menganalisis perubahan nilai loss dan akurasi selama proses training berlangsung. Gambar 6. menunjukkan kurva perkembangan performa model ReMixMatch-DenseNet169 pada setiap *epoch*, yang terdiri dari perbandingan nilai *training loss* dan *validation loss* serta perubahan nilai *training accuracy* dan *validation accuracy*. Analisis kurva ini digunakan untuk mengamati kemampuan model dalam mempelajari pola data, kestabilan proses pembelajaran, serta indikasi terjadinya *overfitting* atau *underfitting* selama proses pelatihan yang ditandai dengan adanya kesenjangan yang signifikan antara performa pada data pelatihan dan data validasi, maupun *underfitting*, yang ditandai dengan rendahnya performa model pada kedua subset data tersebut secara bersamaan, sehingga dapat diperoleh pemahaman yang lebih komprehensif mengenai kondisi dan kualitas proses pelatihan model secara keseluruhan.



Gambar 6. Analisis Kurva Pelatihan ReMixMatch-DenseNet169

Berdasarkan Gambar 3(a), nilai loss pada data pelatihan mengalami penurunan secara bertahap hingga mencapai nilai yang lebih rendah, sedangkan nilai *validation loss* juga menunjukkan tren penurunan yang mengindikasikan model mampu melakukan generalisasi terhadap data validasi. Sementara itu, pada Gambar 3(b), peningkatan nilai akurasi pada data pelatihan dan validasi menunjukkan bahwa model mampu meningkatkan kemampuan klasifikasi seiring bertambahnya jumlah *epoch*. Kedekatan antara kurva akurasi pelatihan dan validasi menunjukkan bahwa model memiliki performa yang stabil tanpa mengalami perbedaan generalisasi yang signifikan.

3.3 Rekapitulasi Hasil Pengujian

Pengujian pengaruh jumlah data berlabel terhadap performa model dilakukan menggunakan beberapa skenario jumlah label per kelas, yaitu 10, 20, 30, dan 40 data berlabel untuk setiap kelas. Setiap skenario dievaluasi menggunakan data validasi yang sama sehingga perbandingan performa antar konfigurasi dapat dilakukan secara objektif. Evaluasi performa model dilakukan menggunakan beberapa metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian model ReMixMatch-DenseNet169 pada setiap skenario jumlah data berlabel ditampilkan pada Tabel 1.

Tabel 1. Rekapitulasi Performa Model ReMixMatch-DenseNet169 pada Seluruh Skenario Eksperimen

Skenario (Label/Kelas)	Accuracy	Precision	recall	F1-Score
10 label per kelas	85,25%	86,14%	85,47%	85,61%
20 label per kelas	91,33%	91,87%	91,45%	91,52%
30 label per kelas	95,11%	92,45%	92,18%	92,26%
40 label per kelas	96,43%	96,47%	96,43%	96,42%

Berdasarkan hasil pengujian pada Tabel 1, peningkatan jumlah data berlabel memberikan pengaruh terhadap kemampuan model dalam melakukan proses klasifikasi. Pada skenario dengan jumlah label terbatas, yaitu 10 label per kelas, model memperoleh akurasi sebesar 85,25%. Seiring dengan meningkatnya jumlah data berlabel menjadi 20 dan 30 label per kelas, performa model mengalami peningkatan karena model memperoleh informasi fitur yang lebih banyak selama proses pembelajaran.

Pada skenario 40 label per kelas, model memperoleh performa terbaik dengan nilai *accuracy* sebesar 96,43%, *precision* sebesar 96,47%, *recall* sebesar 96,43%, dan *F1-score* sebesar 96,42%. Hasil tersebut menunjukkan bahwa peningkatan jumlah data berlabel membantu DenseNet169 dalam mempelajari karakteristik visual dari citra Chest X-Ray secara lebih optimal.

Selain memanfaatkan data berlabel, algoritma ReMixMatch juga berperan dalam meningkatkan kemampuan pembelajaran melalui pemanfaatan data tidak berlabel. Mekanisme *pseudo-labeling* dan *consistency regularization* memungkinkan model memperoleh informasi tambahan dari data yang belum memiliki anotasi, sehingga proses pembelajaran tetap dapat berlangsung secara efektif pada kondisi keterbatasan data berlabel.

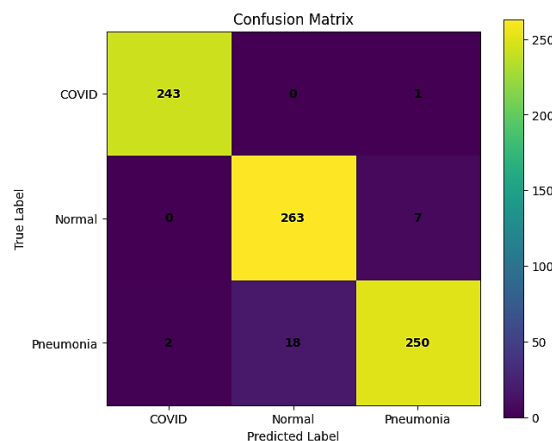
3.4 Analisis Per Kelas Berdasarkan Confusion Matrix

Pola sebaran prediksi model dianalisis secara lebih mendalam melalui confusion matrix yang disajikan pada Gambar 7. Berdasarkan visualisasi tersebut, pada skenario terbaik (40 label per kelas), model mencatatkan hasil sebagai berikut: kelas COVID-19 berhasil diprediksi dengan benar sebanyak 243 dari 244 citra validasi, dengan hanya 1 citra yang salah terklasifikasi sebagai Pneumonia. Selanjutnya, kelas Normal mencatatkan 263 prediksi benar dari 270 citra validasi, dengan 7 citra salah terklasifikasi sebagai Pneumonia. Sementara itu, kelas Pneumonia mencatatkan 250 prediksi benar dari 270 citra validasi, dimana terdapat 18 citra yang salah terklasifikasi sebagai Normal dan 2 citra sebagai COVID-19.

Dari keempat skenario yang diuji dalam penelitian ini, kelas Pneumonia secara konsisten menunjukkan performa yang paling sering mengalami kesalahan prediksi, di mana tingkat kesalahan yang terjadi pada kelas ini juga tampak jauh lebih fluktuatif dibandingkan dengan kelas Normal maupun COVID-19 yang relatif lebih stabil pada setiap skenario pengujian. Fenomena ini kemungkinan besar disebabkan oleh karakteristik visual dari citra itu sendiri,

mengingat pola infiltrat yang menjadi indikator utama pada kasus Pneumonia kerap memiliki kemiripan visual dengan bayangan yang secara normal terdapat pada struktur paru-paru, sehingga batas pembeda antara kedua kelas tersebut menjadi tipis dan sulit dikenali secara visual, baik oleh model maupun secara kasat mata. Temuan ini menjadi pengingat penting bahwa metrik *accuracy* semata tidak cukup untuk dijadikan acuan tunggal dalam menilai kelayakan suatu model di domain medis, sehingga aspek-aspek lain seperti *recall* per kelas serta pola kesalahan yang tampak pada *confusion matrix* juga perlu ikut dicermati secara mendalam, terutama untuk kelas-kelas yang memiliki risiko tinggi apabila terjadi kesalahan atau keterlambatan dalam proses diagnosis (Mosquera et al., 2024).

Evaluasi performa model klasifikasi pada penelitian ini tidak hanya dilakukan berdasarkan nilai akurasi secara keseluruhan, mengingat metrik tersebut belum cukup untuk menggambarkan kemampuan model secara menyeluruh, terutama apabila terdapat ketidakseimbangan jumlah data antar kelas, sehingga diperlukan pula analisis lebih lanjut terhadap distribusi hasil prediksi pada setiap kelas dengan menggunakan *confusion matrix* sebagai alat bantu visualisasi. Gambar 7. menunjukkan *confusion matrix* yang dihasilkan dari model ReMixMatch-DenseNet169 pada data pengujian, yang menggambarkan perbandingan secara rinci antara label sebenarnya (*true label*) dengan label hasil prediksi (*predicted label*) untuk tiga kategori citra Chest X-Ray yang menjadi fokus penelitian ini, yaitu COVID-19, Normal, dan Pneumonia. Melalui visualisasi *confusion matrix* tersebut, dapat diketahui sejauh mana kemampuan model dalam mengenali serta mengklasifikasikan masing-masing kelas dengan tepat, sekaligus dapat diidentifikasi pola-pola kesalahan klasifikasi yang mungkin terjadi antar kategori, sehingga dapat memberikan gambaran yang lebih komprehensif mengenai kelebihan maupun keterbatasan model dalam membedakan karakteristik visual dari masing-masing kelas citra.



Gambar 7. Analisis Distribusi Klasifikasi (Confusion Matrix)

Berdasarkan Gambar 7, model mampu melakukan klasifikasi dengan baik pada ketiga kelas yang diuji. Pada kelas COVID-19, sebanyak 243 citra berhasil diklasifikasikan dengan benar dan hanya terdapat 1 citra yang salah diprediksi sebagai Pneumonia. Pada kelas Normal, model berhasil mengenali 263 citra dengan benar, sedangkan 7 citra mengalami kesalahan klasifikasi sebagai Pneumonia. Sementara itu, pada kelas Pneumonia, sebanyak 250 citra berhasil diklasifikasikan secara tepat, dengan beberapa kesalahan prediksi terhadap kelas COVID-19 dan Normal. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan karakteristik visual antar kelas, meskipun masih terdapat beberapa kemiripan pola citra terutama antara kategori Pneumonia dan Normal.

3.5 Analisis Klasifikasi Per Kelas (Classification Report)

Kemampuan model dalam melakukan klasifikasi pada setiap kelas penyakit dianalisis lebih lanjut menggunakan *classification report* berdasarkan hasil prediksi pada skenario terbaik yang diperoleh dalam penelitian ini, yaitu skenario penggunaan 40 data berlabel pada setiap kelas. Evaluasi menggunakan *classification report* dilakukan untuk mengetahui performa model secara lebih spesifik pada masing-masing kategori citra Chest X-Ray. Parameter evaluasi yang digunakan meliputi *precision*, *recall*, dan *F1-Score*. Ketiga metrik tersebut memberikan informasi mengenai kemampuan model dalam menghasilkan prediksi yang tepat, kemampuan model dalam mengenali seluruh data pada suatu kelas, serta keseimbangan antara *precision* dan *recall*.

Tabel 2. Classification Report Per Kelas Penyakit Paru-Paru (Skenario 40 Label Per Kelas)

Kelas	Precision	Recall	F1-Score
COVID-19	99,18%	99,59%	99,38%
Normal	93,59%	97,41%	95,46%
Pneumonia	96,90%	92,59%	94,69%

Berdasarkan Tabel 2, kelas COVID-19 memperoleh performa klasifikasi tertinggi dengan nilai *precision* sebesar 99,18%, *recall* sebesar 99,59%, dan *F1-Score* sebesar 99,38%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali citra COVID-19. Berdasarkan *confusion matrix*, sebagian besar citra

COVID-19 berhasil diklasifikasikan dengan benar, dimana hanya terdapat satu citra yang mengalami kesalahan prediksi sebagai Pneumonia.

Pada kelas Normal, model memperoleh nilai *precision* sebesar 93,59%, *recall* sebesar 97,41%, dan *F1-Score* sebesar 95,46%. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar citra dengan kondisi Normal. Namun, masih terdapat beberapa kesalahan klasifikasi yang terjadi karena adanya kemiripan karakteristik visual antara citra Normal dengan kelas Pneumonia. Berdasarkan *confusion matrix*, sebanyak tujuh citra Normal mengalami kesalahan prediksi sebagai Pneumonia.

Sementara itu, kelas Pneumonia memperoleh nilai *precision* sebesar 96,90%, *recall* sebesar 92,59%, dan *F1-Score* sebesar 94,69%. Nilai *recall* yang lebih rendah dibandingkan kelas lainnya menunjukkan bahwa masih terdapat beberapa citra Pneumonia yang belum berhasil dikenali secara tepat oleh model. Berdasarkan *confusion matrix*, sebagian besar kesalahan klasifikasi terjadi ketika citra Pneumonia diprediksi sebagai Normal, yaitu sebanyak 18 citra, serta terdapat dua citra yang diprediksi sebagai COVID-19.

3.6 Analisis Interpretabilitas Visual Menggunakan Grad-CAM

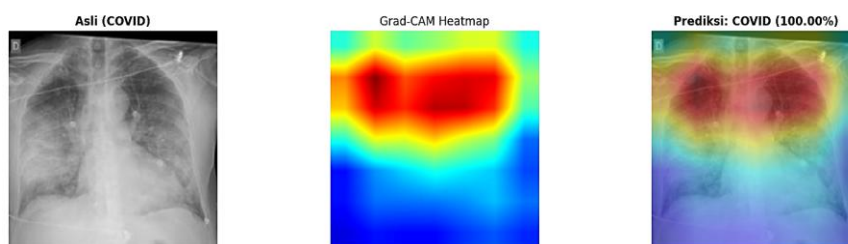
Selain dilakukan evaluasi berdasarkan metrik kuantitatif seperti *precision*, *recall*, dan *F1-Score*, penelitian ini juga melakukan analisis interpretabilitas model dengan menggunakan metode Grad-CAM (*Gradient-weighted Class Activation Mapping*), mengingat evaluasi yang hanya berdasarkan angka semata belum cukup untuk memahami bagaimana model sebenarnya mengambil keputusan dalam melakukan klasifikasi terhadap suatu citra. Analisis interpretabilitas ini bertujuan untuk mengetahui area visual tertentu pada citra Chest X-Ray yang memiliki kontribusi paling besar terhadap keputusan klasifikasi yang dihasilkan oleh model DenseNet169, sehingga dapat diketahui bagian-bagian citra mana yang menjadi fokus utama perhatian model dalam membedakan satu kelas dengan kelas lainnya. Penggunaan metode Grad-CAM ini membantu memberikan gambaran yang lebih jelas mengenai karakteristik fitur yang dipelajari oleh model selama proses pelatihan berlangsung, sekaligus berperan dalam mengurangi sifat black-box yang selama ini menjadi salah satu keterbatasan utama pada pendekatan deep learning, di mana proses pengambilan keputusan oleh model seringkali sulit dijelaskan secara transparan (Kamal et al., 2022).

Pada hasil visualisasi Grad-CAM yang diperoleh, area dengan intensitas warna merah dan kuning menunjukkan bagian citra yang memiliki kontribusi aktivasi paling tinggi terhadap hasil prediksi yang dihasilkan oleh model, yang berarti bagian-bagian tersebut menjadi fokus perhatian utama model dalam menentukan keputusan klasifikasi, sedangkan area yang berwarna biru menunjukkan tingkat kontribusi aktivasi yang lebih rendah, yang mengindikasikan bahwa bagian tersebut tidak terlalu berpengaruh secara signifikan terhadap proses pengambilan keputusan oleh model. Berdasarkan hasil visualisasi yang telah dilakukan terhadap berbagai sampel citra dari masing-masing kelas, dapat diamati bahwa model menunjukkan pola perhatian yang berbeda-beda pada setiap kelas citra Chest X-Ray, di mana perbedaan pola aktivasi ini turut mengindikasikan bahwa model telah mempelajari karakteristik visual yang berbeda pula dalam membedakan satu kategori penyakit dengan kategori lainnya. Pada citra COVID-19 (Gambar 8(a)), area aktivasi model terlihat berada pada beberapa bagian wilayah paru yang memiliki perubahan karakteristik visual. Hal ini menunjukkan bahwa model memanfaatkan informasi tekstur dan pola intensitas pada area tertentu dalam citra untuk menghasilkan prediksi kelas COVID-19.

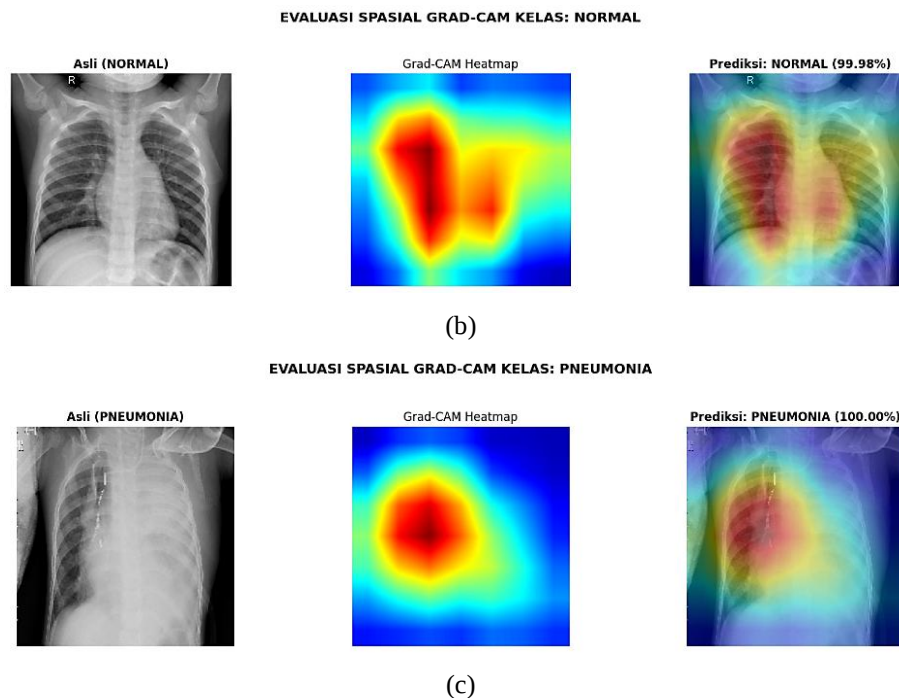
Pada citra Pneumonia (Gambar 8(b)), aktivasi model terlihat lebih terkonsentrasi pada area tertentu dibandingkan kelas lainnya. Pola tersebut menunjukkan bahwa model memberikan perhatian terhadap bagian citra yang memiliki karakteristik visual yang berbeda dari kondisi normal. Sementara itu, pada citra Normal (Gambar 8(c)), distribusi aktivasi model terlihat lebih menyebar dan tidak menunjukkan satu area dominan tertentu. Hal ini sesuai dengan kondisi citra normal yang tidak memiliki pola kelainan visual yang spesifik.

Selain melakukan evaluasi berdasarkan metrik performa klasifikasi, penelitian ini juga melakukan analisis interpretabilitas model menggunakan metode (Grad-CAM). Analisis ini bertujuan untuk mengetahui bagian atau area tertentu pada citra Chest X-Ray yang menjadi perhatian model dalam mengambil keputusan klasifikasi. Gambar 8. menunjukkan hasil visualisasi Grad-CAM pada masing-masing kelas, yaitu COVID-19, Normal, dan Pneumonia, yang terdiri dari citra asli, peta aktivasi (heatmap), serta hasil prediksi dengan area perhatian model yang ditampilkan melalui *overlay*.

EVALUASI SPASIAL GRAD-CAM KELAS: COVID



(a)



Gambar 8. Visualisasi Grad-CAM pada Tiga Kelas Citra Chest X-Ray

Berdasarkan hasil visualisasi Grad-CAM, model ReMixMatch-DenseNet169 mampu memfokuskan perhatian pada area paru-paru yang relevan dalam proses klasifikasi. Pada kelas COVID-19, aktivasi model terlihat terkonsentrasi pada area paru yang mengalami perubahan pola visual akibat infeksi. Pada kelas Pneumonia, perhatian model lebih terarah pada area dengan indikasi kelainan jaringan paru. Sementara itu, pada kelas Normal, model cenderung memperhatikan struktur paru tanpa adanya pola abnormal yang dominan. Hasil visualisasi ini menunjukkan bahwa keputusan model tidak hanya berdasarkan nilai prediksi, tetapi juga didukung oleh representasi fitur visual dari area citra yang memiliki kontribusi terhadap proses klasifikasi.

3.7 Pembahasan

Berdasarkan keseluruhan hasil eksperimen, pendekatan ReMixMatch menunjukkan kemampuan dalam memanfaatkan data tidak berlabel untuk meningkatkan performa klasifikasi citra Chest X-Ray pada kondisi keterbatasan data berlabel. Hasil pengujian menunjukkan bahwa model mampu mencapai performa klasifikasi dengan *accuracy* sebesar 96,43% pada skenario terbaik dengan penggunaan 40 label per kelas. Hasil tersebut menunjukkan bahwa penambahan data berlabel memberikan kontribusi terhadap proses pembelajaran model, sementara mekanisme *semi-supervised learning* pada ReMixMatch tetap memungkinkan model memperoleh informasi tambahan dari data tidak berlabel..

Peningkatan performa model dipengaruhi oleh beberapa komponen utama pada ReMixMatch, yaitu *pseudo-labeling*, *distribution alignment*, *MixUp augmentation*, dan *consistency regularization*. Mekanisme *pseudo-labeling* memungkinkan model menghasilkan label sementara berdasarkan hasil prediksi terhadap data tidak berlabel. Label tersebut kemudian dimanfaatkan sebagai informasi tambahan dalam proses pembelajaran. Selanjutnya, *distribution alignment* berperan dalam menyesuaikan distribusi prediksi agar lebih mendekati distribusi kelas yang sebenarnya sehingga dapat mengurangi kemungkinan terjadinya bias prediksi terhadap kelas tertentu (Sohn et al., 2020).

Selain itu, penerapan *MixUp augmentation* membantu meningkatkan variasi data pelatihan melalui proses kombinasi antar sampel, sehingga model dapat membangun batas keputusan yang lebih baik antar kelas. Sementara itu, *consistency regularization* membantu menjaga kestabilan prediksi model terhadap perubahan citra akibat proses augmentasi. Kombinasi mekanisme tersebut membuat model tetap dapat memanfaatkan informasi dari data tidak berlabel secara efektif meskipun jumlah data anotasi yang tersedia terbatas.

Dibandingkan dengan pendekatan *semi-supervised learning* lain seperti FixMatch, ReMixMatch memiliki mekanisme tambahan berupa *distribution alignment* dan *augmentation anchoring* yang dirancang untuk meningkatkan pemanfaatan data tidak berlabel. Perbedaan tersebut memungkinkan model memperoleh informasi pembelajaran yang lebih luas karena tidak hanya bergantung pada sampel dengan tingkat kepercayaan prediksi yang tinggi.

Hasil penelitian ini menunjukkan bahwa kombinasi ReMixMatch dan DenseNet169 mampu menghasilkan performa klasifikasi yang baik pada citra Chest X-Ray dengan jumlah data berlabel yang terbatas. Hal tersebut menunjukkan bahwa pendekatan *semi-supervised learning* memiliki potensi dalam mengurangi ketergantungan terhadap proses anotasi manual yang membutuhkan waktu dan keterlibatan tenaga ahli, khususnya pada bidang klasifikasi citra medis.



4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan *semi-supervised deep learning* berbasis ReMixMatch dengan arsitektur DenseNet169 mampu menghasilkan performa klasifikasi yang tinggi pada tiga kelas citra Chest X-Ray, yaitu Normal, COVID-19, dan Pneumonia, meskipun menggunakan jumlah data berlabel yang terbatas. Pada skenario terbaik dengan 40 label per kelas, model memperoleh *accuracy* sebesar 96,43%. Berdasarkan hasil evaluasi model, kombinasi ReMixMatch dan DenseNet169 mampu melakukan klasifikasi terhadap masing-masing kelas dengan performa yang baik berdasarkan hasil pengujian menggunakan *precision*, *recall*, dan *F1-Score*. Pada kondisi keterbatasan anotasi, model tetap mampu memanfaatkan informasi dari data tidak berlabel melalui mekanisme *pseudo-labeling*, *distribution alignment*, *MixUp augmentation*, dan *consistency regularization* sehingga dapat membantu meningkatkan kemampuan generalisasi model. Analisis menggunakan *confusion matrix* menunjukkan bahwa sebagian besar prediksi berhasil dilakukan dengan benar, meskipun kesalahan klasifikasi masih ditemukan terutama pada kelas Pneumonia yang memiliki kemiripan karakteristik visual dengan kelas Normal. Selain itu, visualisasi Grad-CAM memberikan informasi mengenai area citra Chest X-Ray yang berkontribusi terhadap keputusan klasifikasi model sehingga membantu meningkatkan interpretabilitas terhadap proses pengambilan keputusan jaringan deep learning. Temuan tersebut memberikan kontribusi terhadap pengembangan metode klasifikasi citra Chest X-Ray melalui penerapan algoritma semi-supervised learning ReMixMatch yang dipadukan dengan arsitektur DenseNet169 pada kondisi keterbatasan data berlabel. Pendekatan yang diusulkan menunjukkan bahwa pemanfaatan data berlabel dan tidak berlabel secara bersamaan dapat menghasilkan model klasifikasi yang andal tanpa bergantung pada ketersediaan data berlabel dalam jumlah besar. Selain itu, integrasi analisis interpretabilitas menggunakan Grad-CAM memberikan penjelasan visual terhadap dasar pengambilan keputusan model, sehingga mendukung pengembangan sistem Computer-Aided Diagnosis (CAD) yang lebih transparan pada bidang pencitraan medis. Penelitian ini masih memiliki ruang pengembangan karena keterbatasannya yang hanya menguji satu arsitektur backbone DenseNet169 dan satu sumber dataset publik. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi penggunaan dataset multi-sumber guna menguji ketahanan model terhadap variasi citra dari alat rontgen yang berbeda, menguji arsitektur backbone mutakhir lainnya atau melakukan studi komparatif mendalam dengan metode *semi-supervised learning* sejenis seperti FixMatch, SoftMatch, atau FlexMatch, dan melibatkan proses validasi klinis secara langsung bersama tenaga ahli radiologi guna mengevaluasi tingkat kesiapan dan kelayakan model jika diimplementasikan pada lingkungan medis dunia nyata.

REFERENCES

- Ahmed, I. A., Senan, E. M., Salameh, H., Shatnawi, A., Alkhraisha, Z. M., & Al-azzam, M. M. A. (2023). Multi-Techniques for Analyzing X-ray Images for Early Detection and Differentiation of Pneumonia and Tuberculosis Based on Hybrid Features. *Diagnostics*, 13(4), 814. <https://doi.org/10.3390/diagnostics13040814>
- Cohen, J. P., Morrison, P., Roth, K., Dao, L., Duong, T. Q., & Ghasseni, M. (2020). *COVID-19 Image Data Collection : Prospective Predictions are the Future*. 1–38.
- Delisle, P., Anctil-robaille, B., Desrosiers, C., & Lombaert, H. (2021). Realistic Image Normalization for Multi-Domain Segmentation. *Medical Image Analysis*, 74. <https://doi.org/10.1016/j.media.2021.102191>
- Finaldi, F. (2024). Gambaran Foto Toraks dan Distribusi Bakteri Patogen pada Pasien Pneumonia Chest X-Ray Overview and Distribution of Pathogenic Bacteria in Pneumonia. *Jurnal MedScientiae*, 3(2), 198–205. <https://doi.org/10.36452/jmedscientiae.v3i2.3144>
- Galbusera, F., & Cina, A. (2024). Image annotation and curation in radiology: an overview for machine learning practitioners. *European Radiology Experimental*, 8(1), 11. <https://doi.org/10.1186/s41747-023-00408-y>
- Haffandi, M. Y., Haerani, E., Syafria, F., & Oktavia, L. (2022). Klasifikasi penyakit paru-paru dengan menggunakan metode naïve bayes classifier. *TEKINKOM*, 5, 176–186. <https://doi.org/10.37600/tekinkom.v5i2.649>
- Hammoudi, K. (2021). Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19. *Journal of Medical Systems*, 45(7), 75. <https://doi.org/10.1007/s10916-021-01745-4>
- Huang, G., Weinberger, K. Q., Zhuang, L., & Lauren, van der M. (2017). *Densely Connected Convolutional Networks*. 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
- Kamal, U., Zunaed, M., Nizam, N. B., Hasan, T., & Member, S. (2022). Anatomy-XNet: An Anatomy Aware Convolutional Neural Network for Thoracic Disease Classification in Chest X-rays. *IEEE Journal of Biomedical and Health Informatics*, 26(11), 5518–5528. <https://doi.org/10.1109/JBHI.2022.3199594>
- Kim, H. E., Linan, A. C., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification : a literature review. *BMC Medical Imaging*, 22, 69. <https://doi.org/10.1186/s12880-022-00793-7>
- Liu, J., Parnell, C., & Summers, R. M. (2021). *Self and Mixed Supervision to Improve Training Labels for Multi- Class Medical Image Segmentation*. 116–123. https://doi.org/10.1007/978-3-030-87237-3_12
- Mazumder, A., Chakraborty, A., Datta, A., Choudhuri, S., Wadadekar, Y., Nirupam, R., & C.H., I.-C. (2020). Characterizing EoR foregrounds : A study of the Lockman Hole Region at 325 MHz. *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 495(4), 4071–4084.
- Mosquera, C., Ferrer, L., Milone, D. H., Luna, D., & Ferrante, E. (2024). *Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance*. 34(12), 7895–



7903. <https://doi.org/10.1007/s00330-024-10834-0>
- Mousavi, M., & Hosseini, S. (2024). A deep convolutional neural network approach using medical image classification. *BMC Medical Informatics and Decision Making*, 24(1), 239. <https://doi.org/10.1186/s12911-024-02646-5>
- Oltu, B., Güney, S., Yuksel, S. E., & Dengiz, B. (2025). Automated classification of chest X-rays : a deep learning approach with attention mechanisms. *BMC Medical Imaging*, 25(1), 32. <https://doi.org/10.1186/s12880-025-01604-5>
- Rahimzadeh, M., & Attar, A. (2020). Detecting Covid-19 And P Neumonia From Chest X-Ray Images Based On The Concatenation Of Xception And Res Net 50V2. *Informatics in Medicine Unlocked*, 19, 100360. <https://doi.org/10.1016/j.imu.2020.100360>
- Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. (2020). Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. *IEEE Access*, 8, 172610–172619. <https://doi.org/10.1109/ACCESS.2020.3020527>
- Shakya, K. S., Alavi, A., Porteous, J., Priti, K., Laddi, A., & Jaiswal, M. (2024). A Critical Analysis of Deep Semi-Supervised Learning Approaches for Enhanced Medical Image Classification. *Applied Sciences*, 14(11), 4788. <https://doi.org/10.3390/app14114788>
- Sohn, K., Berthelot, D., Carlini, N., Cubuk, E. D., Kurakin, A., Zhang, H., & Raffel, C. (2020). *ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring*. <https://doi.org/10.48550/arXiv.1911.09785>
- Wakili, M. A., Shehu, H. A., Sharif, H., Sharif, H. U., Umar, A., Kusetogullari, H., Ince, I. F., & Uyaver, S. (2022). *Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning*. 2022(1). <https://doi.org/10.1155/2022/8904768>