

KLASIFIKASI PENYAKIT PNEUMONIA PADA X-RAY PARU-PARU MENGGUNAKAN MODEL HYBRID GRAY LEVEL CO-OCCURRENCE MATRIX DAN ARTIFICIAL NEURAL NETWORK

(Pneumonia Disease Classification on Chest X-ray Images Using a Hybrid Model of Gray Level Co-occurrence Matrix and Artificial Neural Network)

Amdila Rahmadi*^[1], I Gede Pasek Suta Wijaya^[1], Pahrul Irfan^[1]

^[1]Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: amdilarahmadi@gmail.com, gpsutawijaya@unram.ac.id, irfan@staff.unram.ac.id

Abstract

Pneumonia is a leading cause of morbidity and mortality, particularly in children, requiring rapid and accurate diagnosis. This study proposes a hybrid classification model that combines Gray Level Co-occurrence Matrix (GLCM) texture feature extraction with an Artificial Neural Network (ANN) to analyze chest X-ray images. The dataset consisted of 3,150 images, balanced using random undersampling. GLCM features were extracted across multiple distances and four orientations, generating 19 texture features per image. Seven experimental scenarios were conducted to evaluate ANN architectures with 2, 3, and 4 fully connected layers to identify the most effective configuration. The best-performing model achieved an accuracy of 91.50%, with precision, recall, and F1-score of 0.91, demonstrating consistent performance in distinguishing normal and pneumonia cases. Due to its relatively low computational complexity, this approach is suitable for low-resource healthcare settings. Future work will focus on expanding the dataset and validating the model with clinical data to enhance real-world applicability.

Keywords: image classification, ANN, GLCM, pneumonia, X-ray, hybrid model

*Corresponding Author

I. PENDAHULUAN

Pneumonia merupakan salah satu penyebab utama kematian dan morbiditas pada anak-anak di seluruh dunia [1]. Data Kementerian Kesehatan Republik Indonesia mencatat *pneumonia* menempati peringkat kedua sebagai penyebab kematian balita [2]. Selain dampak kesehatan, penyakit ini juga memberikan beban ekonomi dan sosial yang cukup besar bagi masyarakat dan negara.

Deteksi dini dan penanganan cepat menjadi kunci utama dalam menurunkan angka kematian akibat *pneumonia*. Namun, berbagai kendala masih dihadapi, seperti kurangnya pemanfaatan teknologi dalam sistem pelayanan kesehatan [3]. Kondisi ini dapat menyebabkan proses diagnosis berjalan lebih lambat dan membuat pelayanan kesehatan kurang optimal.

Peluang penggunaan AI di bidang kesehatan dapat diwujudkan melalui pengembangan mesin diagnosis *pneumonia*. Tantangan seperti gejala awal yang tidak spesifik, minimnya radiolog di banyak

daerah, serta prosedur analisis gambar dapat memperlambat proses diagnosis. Keterlambatan dalam diagnosis dapat meningkatkan beban kerja tenaga medis dan fasilitas kesehatan, yang akhirnya mengurangi efisiensi sistem kesehatan secara keseluruhan.

Teknologi kecerdasan buatan (AI) dapat membantu proses diagnosis secara cepat. Salah satu pendekatan yang berkembang adalah penggunaan metode *Machine Learning* dan *Deep Learning*. Penerapan *machine learning* pada gambar x-ray paru-paru dapat digunakan untuk proses awal diagnosis penyakit ini sehingga dapat membantu tenaga medis dalam mengambil keputusan awal.

Sejumlah upaya telah dilakukan dalam rangka mengembangkan sistem diagnosis berbasis AI untuk *pneumonia*, baik dengan metode *machine learning* maupun *deep learning*. Namun, tantangan selanjutnya seperti keterbatasan perangkat yang kurang memadai khususnya di fasilitas Kesehatan seperti puskesmas atau klinik daerah. Sebagian model

terutama *deep learning* memerlukan sumber daya komputasi yang tinggi dalam hal kapasitas memori, penyimpanan dan kecepatan pemrosesan data.

Sebagai solusi atas permasalahan tersebut, penulis mengangkat judul “Klasifikasi Penyakit *Pneumonia* pada *X-ray* Paru-Paru Menggunakan Model *Hybrid* GLCM-ANN.” Penelitian ini bertujuan untuk mengembangkan model AI sebagai diagnosis penyakit *pneumonia* yang menggunakan metode *hybrid* lainnya dengan menggabungkan metode GLCM untuk ekstraksi fitur tekstur dan *Layer Fully Connected* pada ANN sebagai metode klasifikasi. Model ini memiliki keunggulan dalam hal jumlah parameter yang jauh lebih sedikit dibandingkan model *deep learning* yang biasanya memiliki jutaan parameter. Dengan komputasi yang lebih ringan, pendekatan ini diharapkan mampu menghasilkan sistem diagnosis yang dapat diimplementasikan sebagai alternatif metode yang mampu mendeteksi penyakit *pneumonia* secara lebih efektif di Indonesia.

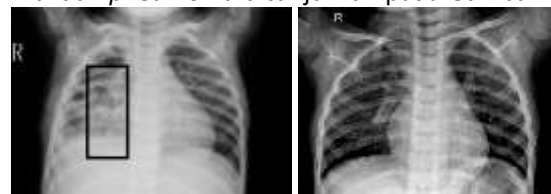
II. TINJAUAN PUSTAKA

Penelitian sebelumnya menerapkan beberapa metode yang serupa, salah satunya penerapan model *Hybrid* GLCM-CNN untuk mengidentifikasi dan mengklasifikasi tingkatan dari tumor otak. Penggunaan model ini digunakan pada area tumor pada citra otak yang telah disegmentasi. Sebanyak 260 citra MRI dilatih dengan rincian sebanyak 156 citra yang tingkat rendah (*benign*) dan 104 citra yang tingkat tinggi (*malignant*). Dihasilkan 1560 total fitur dari tersebut yang dibagi menjadi 6 fitur per citra. Ekstraksi fitur GLCM dilakukan pada sudut 0° dan 90°. Adapun fitur GLCM yang digunakan diantaranya fitur *contrast*, *entropy*, *mean* dan *homogeneity*. Fitur tersebut akan digunakan sebagai input untuk lapisan *fully connected* untuk mengklasifikasi tingkatan tumor pada citra otak. *Classification rate* dari citra normal dan tumor mencapai 99.5% pada epoch ke-10 [4].

Penelitian lainnya menggunakan metode yang sama dengan tambahan metode CNN dan *Hybrid* GLCM-SVM sebagai pembanding dalam pengklasifikasian daun tanaman herbal. Sebanyak 480 total dataset berukuran 600×600 pixels yang dibagi rata menjadi 120 citra dalam dua kelas (kelor dan katuk) dan dua skenario (terang dan gelap). Fitur GLCM yang digunakan pada citra diantaranya *energy*, *contrast*, *correlation*, *homogeneity*, *dissimilarity*, dan *ASM*. Hasil GLCM akan dijadikan input pada model SVM dan CNN. Training model dilakukan rasio data sebesar 80:20 atau dengan kata lain 192 citra training dan 48 citra testing. Skenario dilakukan sebanyak 20

epoch dan batch size sebesar 22 pada model CNN. Untuk model SVM dilakukan pengujian dengan empat kernel yang berbeda. Pengujian model mendapatkan hasil akurasi terbesar yaitu 98% pada skenario terang untuk model CNN diikuti dengan model GLCM-SVM dengan kernel linear sebesar 94% dan GLCM-CNN sebesar 88%. Pengujian dengan skenario gelap mendapatkan hasil akurasi terbesar yang sama untuk ketiga model CNN, GLCM-CNN dan GLCM-SVM yaitu 96% [5].

Terdapat penelitian lainnya yang menggunakan dataset *X-ray* paru-paru *pneumonia* dengan metode CNN ataupun metode lain. Secara umum, ciri radiologis *pneumonia* ditandai dengan adanya *infiltrates* atau opasifikasi paru yang baru muncul pada citra *X-ray* [6]. Perbedaan visual dari paru-paru normal dan *pneumonia* ditunjukkan pada Gambar 1.



Gambar 1. Citra *X-ray* Paru-Paru *Pneumonia* (kiri), Normal (kanan)

Berdasarkan radiografi dada standar (CXR), *infiltrates* atau suatu zat padat yang ada pada paru-paru biasanya tampak sebagai area bercak-bercak yang menunjukkan akumulasi cairan, *neutrofil*, atau jaringan inflamasi di *alveoli* [6]. Zat tersebut menyebabkan penderita mengalami gejala nyeri dada dan batuk. Berdasarkan Gambar 1, bercak tersebut dapat dilihat dengan adanya penambahan *pixel* yang berwarna putih keabuan yang mengumpul di satu tempat, seperti pada Gambar kiri dari Gambar 1. Gambar ini memperlihatkan area pada kotak hitam sebagai *infiltrates* yang mengindikasikan paru-paru tersebut mengidap penyakit *pneumonia*. Sedangkan untuk gambar kanan, citra paru-paru tampak lebih bersih dibandingkan gambar kiri, hal ini mengindikasikan paru-paru dalam keadaan normal.

Keakuratan diagnosis *pneumonia* sangat bergantung pada kemampuan radiolog dalam menginterpretasikan citra dengan tepat, hal ini sering dipengaruhi oleh kualitas citra dan pengalaman ahli sebagai radiolog [6]. Dalam beberapa kasus, citra *X-ray* menunjukkan kemiripan antar berbagai penyakit paru, seperti *tuberkulosis*, kanker paru-paru, atau *pneumothorax*, sehingga meningkatkan risiko kesalahan diagnosis [7].

Salah satu penelitian menggunakan model CNN untuk citra *X-ray pneumonia*. Penelitian ini dilakukan pada tahun 2020 yang bertujuan untuk menguji

kinerja *deep learning* pada citra *X-ray* ini menggunakan 5.840 data citra yang terdiri dari 1.575 citra normal dan 4.265 citra *pneumonia*. Penggunaan fungsi ReLU sebagai optimasi, Adam sebagai optimasi dan dengan jumlah *epoch* yang dicoba sebanyak 200, menghasilkan rata-rata akurasi sebesar 89,58% dan rata-rata *loss* sebesar 47,43%. Namun terdapat gap antara hasil pengujian dan pelatihan untuk akurasi sebesar 8,72% dan *loss* 42,97% yang mengakibatkan kasus *overfitting*. Kondisi dataset, formulasi pada model dan jumlah epoch dapat menjadi faktor yang menyebabkan kasus ini terjadi [7].

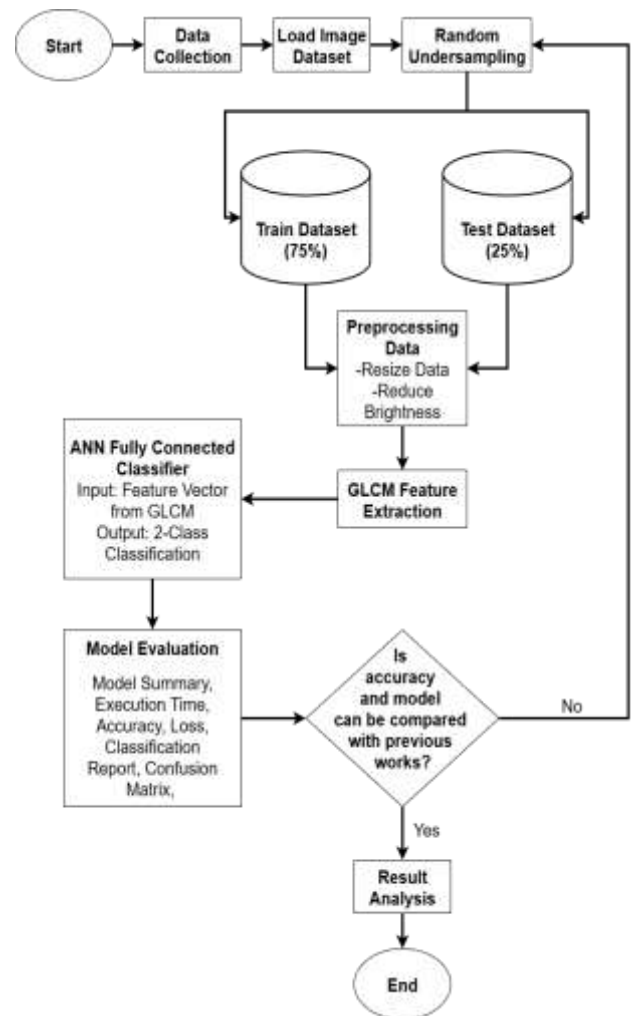
Penelitian lainnya dilakukan pada tahun 2022 mengkaji penerapan metode transfer learning dengan model ResNet152V2 untuk klasifikasi citra *X-ray* paru-paru *pneumonia*. Memanfaatkan dataset yang berisi 3.590 citra, penelitian ini berhasil mencapai akurasi sebesar 91,4% dan nilai *loss* 0,18 setelah melakukan evaluasi sebanyak 20 kali. Model pre-trained dari ResNet152V2 memungkinkan ekstraksi fitur yang efektif dari citra medis. Hasil pelatihan yang tidak mengalami kasus *overfitting* merupakan keunggulan utama dari pendekatan ini [8].

Terdapat beragam penelitian lainnya yang menggunakan pendekatan *machine learning* dan *deep learning* pada citra *X-ray* paru-paru [9], [10], [11]. Penggunaan model dan dataset yang beragam, kasus diagnosa penyakit yang berbeda hingga perbandingan antara model menjadi fokus utama dalam penelitian-penelitian tersebut.

Berdasarkan penelitian sebelumnya, pendekatan yang dilakukan dengan model *hybrid* mendapatkan hasil yang tidak kalah baiknya dengan pendekatan dengan metode *deep learning* maupun *machine learning* murni seperti CNN. Pendekatan *hybrid* ini telah menunjukkan hasil yang positif dalam berbagai kasus klasifikasi citra medis, seperti deteksi tumor otak, pengenalan daun herbal, hingga klasifikasi *pneumonia* dari citra *X-ray* paru-paru. Sehingga, dengan mempertimbangkan tantangan implementasi metode *deep learning* yang cenderung membutuhkan sumber daya komputasi tinggi dan rentan terhadap *overfitting*, dibutuhkan alternatif metode yang lebih ringan namun tetap akurat dalam perancangan klasifikasi penyakit terhadap citra *X-ray*.

III. METODE PENELITIAN

Penelitian ini dilakukan dengan skenario seperti yang dicantumkan pada Gambar 2.

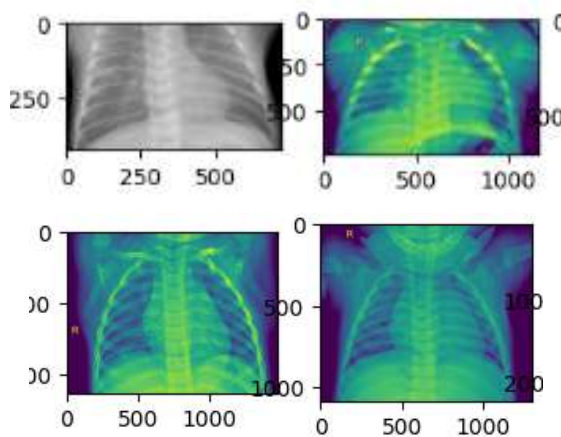


Gambar 2. Skenario Pembuatan dan Pelatihan Model Klasifikasi.

Proses dimulai dari tahap *load* data dan *random sampling* dilanjutkan dengan *pre-processing* data. Kemudian akan dilakukan proses ekstraksi fitur untuk mendapatkan fitur dari data yang akan digunakan pada proses pemodelan dan training. Penjelasan dari setiap proses ini akan dijelaskan pada sub bagian selanjutnya.

3.1. Dataset

Dataset terdiri dari 5.856 citra *X-ray* paru-paru anak-anak, dengan 5.232 citra untuk training dan 624 citra untuk testing. Citra telah melalui proses pelabelan dari dua ahli radiolog dan seorang ahli radiolog senior sebagai pemberi keputusan akhir pada data citra untuk memastikan keakuratan data dan menghindari ketidaksepakatan antara dua ahli lainnya. Penulis menyajikan sampel citra dari dataset sebagai Gambar 3.



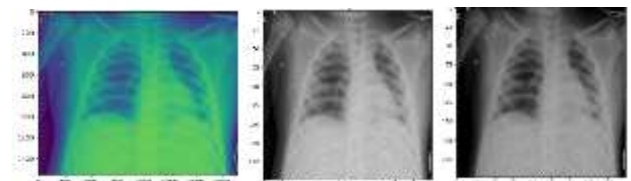
Gambar 3. Sampel Gambar Dataset

Gambar 3 menunjukkan beberapa contoh citra X-ray paru-paru dari dataset, yang mencakup kondisi paru-paru normal(bawah) dan yang terinfeksi *pneumonia*(atas). Citra ini memberikan gambaran visual mengenai variasi struktur paru-paru serta perbedaan karakteristik antara citra normal dan citra dengan indikasi *pneumonia*. Dataset ini disimpan dalam *database* di *mendeley*, sehingga memudahkan peneliti lain untuk mengakses dan memanfaatkan data tersebut dalam pengembangan teknologi untuk diagnosis medis [12].

3.2. Sampling dan Pre-processing

Pada penelitian ini, dilakukan *random undersampling* terhadap kelas mayoritas untuk menciptakan distribusi kelas yang seimbang, dalam hal ini data untuk kelas normal dan *pneumonia* disamakan. Teknik ini dilakukan dengan mengambil sejumlah sampel secara acak dari data berlabel *pneumonia* agar jumlahnya setara dengan kelas minoritas, yaitu data berlabel normal [13]. Jumlah total citra berlabel *pneumonia* adalah 4.265, sedangkan citra berlabel normal berjumlah 1.575. Oleh karena itu, diputuskan untuk menggunakan 1.575 citra dari masing-masing kelas. Pendekatan ini bertujuan untuk mengatasi permasalahan *class imbalance* yang dapat menyebabkan bias pada model.

Pemilihan sampel dilakukan secara acak pada masing-masing kelas dan label dengan rasio sebesar 1181:394 atau 75:25, kemudian dilakukan proses *pre-processing* dengan tiga tahap yaitu mengonversi citra menjadi citra *grayscale*, mengurangi ukuran citra dan menyeragamkannya, dan mengurangi kecerahan citra sebesar 30. Gambar 4 merupakan perbandingan citra sebelum dan sesudah dari tahap ini.



Gambar 4. Citra Original (kiri), Citra setelah *Resizing* (tengah), dan Citra setelah digelapkan (kanan)

Konversi citra menjadi citra *grayscale* dilakukan untuk mendukung proses ekstraksi fitur GLCM dan proses pengurangan intensitas citra agar citra tampak lebih gelap. Pengurangan ukuran citra bertujuan untuk meningkatkan efisiensi komputasi dan kinerja model pembelajaran mesin. Pengurangan citra membuat data menjadi homogen dan mempermudah proses ekstraksi fitur pada citra dikarenakan penggunaan *resizing* yang dapat mengurangi noise dan variansi yang tidak relevan dari data asli [9], [10]. Citra asli (gambar kiri) dengan ukuran 1904x1560 *pixels* akan jauh membutuhkan waktu pemrosesan yang lebih lama dibandingkan dengan citra berukuran 224x224 *pixels* (gambar tengah) karena jumlah *pixel* dari citra lebih sedikit. Kemudian, pengurangan intensitas citra sebesar 30 seperti yang ditunjukkan pada gambar kanan untuk mengurangi bercak putih berukuran kecil terutama pada citra paru-paru normal agar saat proses training agar citra tersebut tidak dianggap sebagai citra *pneumonia*.

3.3. Ekstraksi Fitur

Ekstraksi fitur menggunakan GLCM untuk menangkap tekstur dari gambar X-ray. GLCM dilakukan dengan representasi hubungan antara dua pixel pada gambar berdasarkan rentang jarak dan sudut. Terdapat empat sudut yang digunakan dalam penelitian ini yaitu 0°, 45°, 90°, dan 135°. Parameter jarak (*d*) yang digunakan adalah 1, 2, dan 3 untuk memperoleh representasi *co-occurrence* pada beberapa tingkat kedekatan piksel, sehingga informasi tekstur pada skala pendek dapat terekstraksi secara menyeluruh. Pada penelitian ini, Persamaan (1) hingga Persamaan (10) akan digunakan sebagai fitur untuk *training* selanjutnya [14], [15], [16].

$$\text{Contrast: } \sum_{i,j=1}^N P(i,j)(i-j)^2 \quad (1)$$

$$\text{Dissimilarity: } \sum_{i,j=1}^N P(i,j) |i-j| \quad (2)$$

$$\text{Homogeneity: } \sum_{i,j=1}^N \frac{P(i,j)}{(1+(i-j)^2)^2} \quad (3)$$

$$\text{ASM: } \sum_{i,j=1}^N P^2(i,j) \quad (4)$$

$$\text{Energy: } \sqrt{\sum_{i,j=1}^N P^2(i,j)} \quad (5)$$

$$\text{Correlation: } \sum_{i,j=1}^N P(i,j) \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right] \quad (6)$$

$$\text{Mean: } \frac{\sum_{i,j=1}^N i * P(i,j) + \sum_{i,j=1}^N j * P(i,j)}{2} \quad (7)$$

$$Var: \frac{\sum_{i,j=1}^N P(i,j) * (i - \mu_i)^2 + \sum_{i,j=1}^N P(i,j) * (j - \mu_j)^2}{2} \quad (8)$$

$$Std: \sqrt{\frac{\sum_{i,j=1}^N P(i,j) * (i - \mu_i)^2 + \sum_{i,j=1}^N P(i,j) * (j - \mu_j)^2}{2}} \quad (9)$$

$$Entropy: \sum_{i,j=1}^N P(i,j) (-\ln P(i,j)) \quad (10)$$

3.4. Pemodelan dan Training

Model yang digunakan adalah ANN yang hanya memanfaatkan bagian *fully connected*. Bagian ini digunakan untuk mengubah multidimensional array hasil ekstraksi fitur data ke dalam vektor [17], [18]. Arsitekturnya terdiri dari *input layer*, *dropout*, dan *dense layer*. *Input layer* menerima data fitur, *dropout* digunakan untuk mencegah *overfitting* dengan menonaktifkan unit secara acak selama pelatihan, dan *dense layer* berfungsi untuk proses klasifikasi.

Dalam beberapa percobaan, model menggunakan *dense layer* dan jumlah *neuron* yang bervariasi. Percobaan dilakukan dengan membandingkan model dengan 2, 3 dan 4 *layer*, variasi ini digunakan untuk mengevaluasi *layer* yang paling optimal bagi fitur GLCM, sehingga pemilihan arsitektur akhir dapat dilakukan berdasarkan performa terbaik yang diperoleh dari ketiga konfigurasi tersebut.

Fungsi aktivasi yang digunakan berupa ReLU dan Sigmoid. Fungsi ReLU (*Rectified Linear Unit*) seperti yang diperlihatkan pada Persamaan (11) adalah fungsi matematis pada jaringan syaraf tiruan yang meneruskan nilai input ke lapisan berikutnya [19].

$$f(x) = \max(0, x) \quad (11)$$

Fungsi aktivasi *sigmoid* mengubah nilai input menjadi sebuah nilai diantara 0 dan 1 dengan bentuk kurva S. Penggunaan fungsi ini digunakan dalam layer *fully connected* untuk mengonversi hasil akhir ke dalam nilai berupa probabilitas. Fungsi ini berguna dalam pengambilan keputusan pada sistem untuk

menentukan citra paru-paru yang memiliki penyakit *pneumonia* atau normal. Fungsi ini didefinisikan seperti Persamaan (12) di bawah ini.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

3.5. Evaluasi Model

Evaluasi model dilakukan dengan menganalisa *confusion matrix* dari setiap skenario untuk mengetahui lebih detail tentang performa klasifikasi dari setiap model. Struktur umum *confusion matrix* ada pada Tabel I.

TABEL I. Struktur *Confusion Matrix*

	Predicted Positive	Predicted Negative
Actual Positive	True positive	False negative
Actual Negative	False positive	True negative

Penggunaan *matrix* pada Tabel I ini dilakukan untuk melihat seberapa baik model ketika membedakan antara kelas positif dan negatif, serta mengidentifikasi kesalahan prediksi yang terjadi seperti *false positive* (FP) dan *false negative* (FN). Setiap skenario menghasilkan *confusion matrix* yang berbeda, dikarenakan kemampuan klasifikasi pada kondisi *training* yang berbeda-beda [20], [21].

IV. HASIL DAN PEMBAHASAN

Penulis melakukan percobaan dengan variasi pada tahap *sampling*, *pre-processing*, ekstraksi fitur, pemodelan, dan *training*. Setiap skenario dimulai dengan *random sampling* sesuai rasio tertentu, dilanjutkan dengan *pre-processing* berupa *resizing* dan pengurangan kecerahan citra. Fitur GLCM kemudian diekstrak dari setiap sampel dan digunakan sebagai input untuk model ANN. Rincian dari percobaan yang dilakukan ditunjukkan pada Tabel II.

TABEL II. Tabel Setiap Percobaan.

Sce.	Data Sampling		Resize	Total Fitur	Layer	Neuron (per layer)	Fungsi Aktivasi (per layer)	Weight	Patience	Epochs
	Train	Test								
1	1181	394	224 × 224	19	2	128, 1	ReLU, Sigmoid	{0: 1.5, 1: 0.8}	50	100
2	1181	394	224 × 224	19	2	128, 1	ReLU, Sigmoid	{0: 1.5, 1: 0.8}	-	50
3	1181	394	224 × 224	19	3	128, 64, 1	ReLU, ReLU, Sigmoid	{0: 1.5, 1: 0.8}	50	100
4	1181	394	224 × 224	19	3	128, 64, 1	ReLU, ReLU, Sigmoid	{0: 1.5, 1: 0.8}	-	50
5	1181	394	224 × 224	19	4	128, 64, 32, 1	ReLU, ReLU, ReLU, Sigmoid	{0: 1.5, 1: 0.8}	50	100
6	1181	394	224 × 224	19	4	128, 64, 32, 1	ReLU, ReLU, ReLU, Sigmoid	{0: 1.5, 1: 0.8}	-	50

Tabel II menunjukkan rincian dari delapan skenario percobaan yang dilakukan. Pada setiap skenario, data dibagi menjadi 1.181 data latih dan 394 data uji, dengan ukuran citra yang telah di-*resize* menjadi 224×224 piksel. Pemilihan ukuran ini digunakan agar seluruh citra memiliki dimensi seragam dan tetap mempertahankan pola tekstur pada citra *X-ray*, namun tidak terlalu besar sehingga proses ekstraksi fitur tetap efisien. Jumlah total fitur yang digunakan dalam setiap percobaan adalah 19 fitur yang merepresentasikan karakteristik tekstur dari citra.

Model yang digunakan memiliki variasi dalam jumlah layer, jumlah *neuron* per layer, serta fungsi aktivasi. Untuk lapisan tersembunyi, fungsi aktivasi yang digunakan adalah ReLU, sedangkan untuk lapisan *output* digunakan *Sigmoid*.

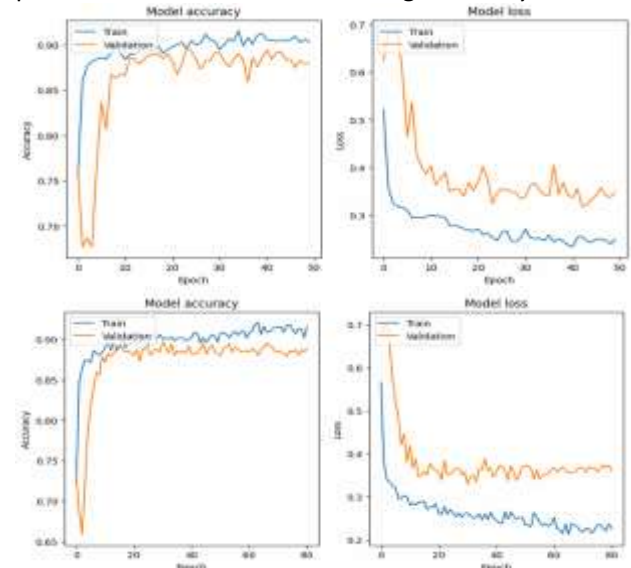
Jumlah *epoch* pada setiap skenario dibuat berbeda. Skenario 1, 3, dan 5 dilatih hingga 100 *epoch* dengan memanfaatkan *callback* (*early stopping*) untuk menghentikan pelatihan lebih awal apabila model tidak menunjukkan peningkatan, sehingga tetap aman dari *overfitting*. Sebaliknya, skenario 2, 4, dan 6 hanya dilatih selama 50 *epoch* tanpa *callback* untuk melihat bagaimana performa model ketika pelatihan dibuat lebih singkat dan tanpa mekanisme penghentian otomatis. Hasil training setiap skenario berada pada kisaran 88% hingga 91% dengan *loss* 26% hingga 32%. Nilai *precision*, *recall*, dan *f1-score* berada di kisaran 88–91%. Tabel III menyajikan rincian akurasi, *loss*, *precision*, *recall*, dan *f1-score*.

TABEL III. Tabel Hasil Training Model

S c e n a r i	Acc	Loss	Precision (macro / weighte d)	Recall (macro/ weighte d)	F1-Score (macro/ weighte d)
1	0.909 9	0.266 1	0.91 / 0.91	0.91 / 0.91	0.91 / 0.91
2	0.898 5	0.273 6	0.90 / 0.90	0.90 / 0.90	0.90 / 0.90
3	0.880 7	0.329 1	0.88 / 0.88	0.88 / 0.88	0.88 / 0.88
4	0.888 3	0.281 7	0.89 / 0.89	0.89 / 0.89	0.89 / 0.89
5	0.897 2	0.297 5	0.90 / 0.90	0.90 / 0.90	0.90 / 0.90
6	0.915 0	0.262 2	0.91 / 0.91	0.91 / 0.91	0.91 / 0.91

Struktur model pada skenario dengan 2 layer cenderung memberikan hasil yang cukup baik.

Skenario 1, yang dilatih selama 100 *epoch*, menghasilkan akurasi 90.99% dengan *loss* 26.61%. Sedangkan skenario 2 yang hanya dilatih selama 50 *epoch* mencatat akurasi sedikit lebih rendah, yaitu 89.85% dengan *loss* 27.36%. Hal ini menunjukkan penambahan jumlah *epoch* dapat mengoptimalkan performa model. Gambar 5 menunjukkan grafik performa dari kedua skenario dengan dua layer.

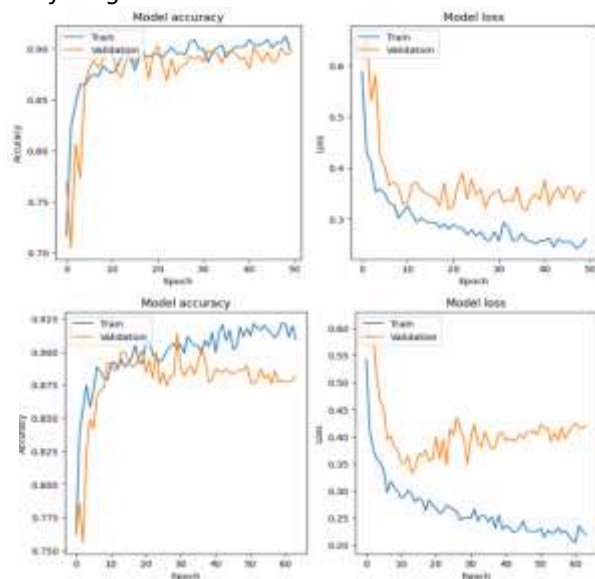


Gambar 5. Grafik Akurasi(kiri) dan Loss(kanan) Skenario 1(atas) dan 2(bawah).

Kedua grafik pada gambar 5 menunjukkan performa dari dua skenario pertama dengan model 2 layer. Performa model pada skenario pertama dengan 50 *epoch* menunjukkan adanya fluktuasi setiap *epoch*-nya. Perbedaan antara *train* dan *validation loss* dari model pertama menunjukkan model mengalami *overfitting*. Sementara itu, skenario kedua dengan 100 *epoch* menunjukkan grafik yang lebih stabil walaupun jarak antara kedua *loss* masih menunjukkan model mengalami *overfitting*. Skenario model dengan 2 layer menunjukkan bahwa penggunaan *epoch* yang lebih banyak dengan tambahan *earlystopping* akan menghasilkan performa yang lebih stabil. Namun, terjadinya *overfitting* pada model ini menunjukkan model kurang mempelajari pola data *validation* dibandingkan data *train*.

Ketika jumlah layer ditingkatkan menjadi 3 layer, hasil yang diperoleh tidak terlalu baik. Skenario 3 yang dilatih selama 100 *epoch* menunjukkan performa terendah dengan akurasi 88.07% dan *loss* tertinggi 32.91%. Sebaliknya, skenario 4 dengan 50 *epoch* mencatat hasil yang lebih stabil dengan akurasi 88.83% dan *loss* 28.17%. Performa model 3 layer seperti yang diperlihatkan pada gambar 6

menunjukkan model masih mengalami fluktuasi dan *overfitting*.

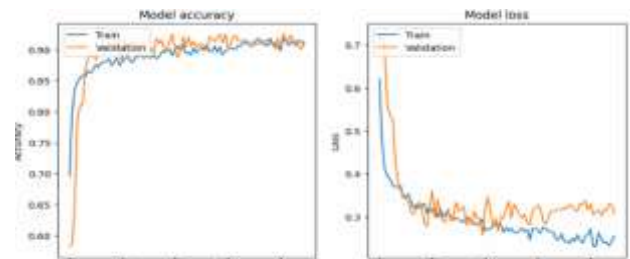
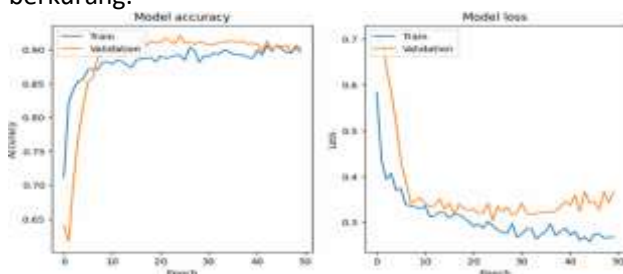


Gambar 6. Grafik Akurasi(kiri) dan *Loss*(kanan)
Skenario 3(atas) dan 4(bawah).

Grafik pada gambar 6 untuk model skenario 3 masih mengalami fluktuasi dan *overfitting*, walaupun dibandingkan dengan skenario sebelumnya jarak antara kedua *loss* mengalami penurunan. Sebaliknya untuk model skenario 4, jarak antara kedua *loss* terlihat cukup besar sehingga model juga mengalami *overfitting*.

Berdasarkan hasil training dan grafik performa untuk model dengan 3 *layer*, dapat disimpulkan bahwa performa model ini lebih kurang dengan data yang digunakan dibandingkan dengan kedua tipe model lainnya. Indikasi terlihat pada hasil *training* yang menurun, grafik yang fluktuatif dan model mengalami *overfitting*.

Skenario 5 dan 6, jumlah *layer* ditingkatkan menjadi 4 *layer* dengan konfigurasi *neuron* 128, 64, 32, dan 1 menghasilkan performa model yang lebih baik. Skenario 6 mencatat performa terbaik dengan akurasi 91.50% dan *loss* 26.22%, diikuti oleh skenario 5 dengan akurasi 90.48% dan *loss* 26.63%. Berdasarkan grafik performa model pada gambar 7, model dengan 4 *layer* menghasilkan hasil yang lebih stabil dan indikasi model untuk mengalami *overfitting* berkurang.



Gambar 7. Grafik Akurasi(kiri) dan *Loss*(kanan)
Skenario 5(atas) dan 6(bawah).

Terlihat untuk grafik pada gambar 7, perbedaan antara kedua *loss* tergolong sedikit sehingga jarak antara kedua *loss* berdekatan. Hal ini menunjukkan model menggeneralisasikan data dengan baik dan tidak mengalami *overfitting*. Peningkatan akurasi dan penurunan *loss* pada grafik pertama dan kedua menunjukkan performa model yang lebih baik dibandingkan dengan model sebelumnya dengan *layer* yang lebih sedikit.

Evaluasi model dari keenam skenario lainnya dapat dilakukan dengan menganalisa *confusion matrix* dari masing-masing skenario. Berdasarkan Tabel IV, model mampu mengklasifikasi data citra x-ray dengan baik. Nilai *true positive* dan *true negative* di setiap skenario menunjukkan performa model berbanding lurus dengan hasil akurasi *training* masing-masing model.

TABEL IV. Tabel *Confusion Matrix*

Sce.	True Positive	True Negative	False Positive	False Negative
1	365	352	42	29
2	346	362	32	48
3	346	354	40	48
4	348	346	48	46
5	356	351	43	38
6	354	356	38	40

Hasil model di setiap skenario menunjukkan performa yang hampir sama walaupun setiap skenario memiliki perbedaan dalam jumlah *layer* dan *epoch*. Skenario dengan model yang mampu mengklasifikasi citra dengan baik ditandai dengan nilai *true positive* dan *true negative* yang seimbang. Berdasarkan tabel IV, model pada skenario 6 dapat mengklasifikasi citra x-ray paru-paru normal dan *pneumonia* dengan seimbang di angka 354 untuk citra normal dan 356 untuk citra *pneumonia*.

Di sisi lain nilai *false positive* dan *false negative* juga berpengaruh dalam kinerja model ketika diaplikasikan di dunia medis. Nilai *false positive* yang mengklasifikasikan citra *pneumonia* sebagai citra normal dapat membahayakan pasien yang seharusnya mendapatkan penanganan menjadi tidak terdeteksi.

Sehingga nilai dari *false positive* harus lebih kecil dari nilai *false negative*. Hal ini menunjukkan bahwa skenario 2, 3 dan 6 menjadi skenario dengan model yang dapat diaplikasikan, namun dalam hal ini skenario 6 tetap menjadi skenario terbaik untuk diaplikasikan dikarenakan akurasi yang diberikan adalah akurasi yang tertinggi dari ketiga skenario tersebut.

V. KESIMPULAN DAN SARAN

Penelitian ini mengembangkan model klasifikasi citra X-ray paru-paru menggunakan pendekatan hybrid yang menggabungkan ekstraksi fitur GLCM dengan arsitektur model ANN. Hasil yang diperoleh menunjukkan bahwa model mampu mengklasifikasikan citra dengan akurasi tertinggi mencapai 91.50%. Selain itu, nilai *precision*, *recall*, dan *f1-score* tertinggi, baik secara makro maupun berbobot, sebesar 0.91, hal ini mengindikasikan model memiliki performa yang seimbang dalam mengenali kedua kelas, yaitu normal dan *pneumonia*. Evaluasi dari masing-masing skenario menunjukkan bahwa beberapa model dapat mengklasifikasikan kedua kelas citra, meskipun pada beberapa konfigurasi terjadi ketidakseimbangan sensitivitas terhadap kedua kelas tersebut.

Pendekatan *hybrid* ini layak dipertimbangkan sebagai metode alternatif dalam tugas klasifikasi citra medis, terutama untuk penggunaan perangkat dengan spesifikasi rendah, dikarenakan jumlah parameter yang dihasilkan relatif sedikit. Untuk pengembangan lebih lanjut, disarankan agar penelitian selanjutnya mempertimbangkan peningkatan jumlah data latih, eksplorasi arsitektur yang lebih kompleks, atau penggantian metode ekstraksi fitur. Selain itu, pengujian lebih lanjut menggunakan data klinis yang lebih bervariasi serta evaluasi langsung dari para ahli radiologi sangat diperlukan untuk memastikan efektivitas model dalam implementasi nyata di dunia medis.

VI. DAFTAR PUSTAKA

- [1] M. A. Khan, A. Bajwa, and S. T. Hussain, "Pneumonia: Recent Updates on Diagnosis and Treatment," Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/microorganisms13030522.
- [2] N. R. Sary, I. Kusumastuti, and R. Sugesti, "Hubungan Pengetahuan Ibu, Sikap, Lingkungan Fisik dan Peran Bidan Dengan Perilaku Ibu dalam Pencegahan Pneumonia pada Balita ISPA di Klinik Kita Depok Jawa Barat Tahun 2023," *SENTRI: Jurnal Riset Ilmiah*, vol. 9, pp. 4257–4271, Aug. 2024.
- [3] C. Kurniawan, D. Sutiningsih, and Martini, "Sistem Aplikasi Berbasis Website dalam Deteksi Dini dan Edukasi Pneumonia," Apr. 2023. [Online]. Available: <http://journal.stikeskendal.ac.id/index.php/PSKM>
- [4] A. Gurunathan and B. Krishnan, "A Hybrid CNN-GLCM Classifier for Detection and Grade Classification of Brain Tumor," Jun. 07, 2021. doi: 10.21203/rs.3.rs-531022/v1.
- [5] P. Purnawansyah *et al.*, "Comparative Study of Herbal Leaves Classification using Hybrid of GLCM-SVM and GLCM-CNN," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 382–389, Aug. 2023, doi: 10.33096/ilkom.v15i2.1759.382-389.
- [6] A. Torres *et al.*, "Pneumonia," Dec. 01, 2021, *Nature Research*. doi: 10.1038/s41572-021-00259-0.
- [7] I. M. Dendi Maysanjaya, "Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network," 2020.
- [8] M. Ramadhan, D. Iskandar Mulyana, and M. Betty Yel, "Optimasi Algoritma CNN Menggunakan Metode Transfer Learning Untuk Klasifikasi Citra X-Ray Paru-Paru Pneumonia dan Non-Pneumonia," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 2, 2022.
- [9] M. R. Hasan and S. M. A. Ullah, "Enhancing Pneumonia Diagnosis: An Ensemble of Deep CNN Architectures for Accurate Chest X-Ray Image Analysis," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2024, pp. 255–269. doi: 10.1007/978-981-99-8937-9_18.
- [10] J. Rakesh, "COVID-19 and Other Pneumonia Diagnosis Using CNN," *Int J Res Appl Sci Eng Technol*, vol. 10, no. 10, pp. 1519–1525, Oct. 2022, doi: 10.22214/ijraset.2022.47249.
- [11] A. Sowmiya, C. Shilaja, G. Nalinashini, and N. Padmavathi, "AI Based COVID Pneumonia Classifier Using Machine Learning," *ECS Trans*, vol. 107, no. 1, pp. 7289–7298, Apr. 2022, doi: 10.1149/10701.7289ecst.
- [12] D. S. Kermay *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
- [13] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," in *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, Institute of Electrical and

- Electronics Engineers Inc., Apr. 2020, pp. 243–248. doi: 10.1109/ICICS49469.2020.239556.
- [14] E. Rianti, I. Fitri, Sumijan, and F. F. Yani, "Development of GLCM Method in Calculate Entropy Value for Digital Visualization in Identifying Childhood Pneumonia Based on Chest X-Ray Images," *International Journal of Online and Biomedical Engineering*, vol. 21, no. 2, pp. 137–156, Feb. 2025, doi: 10.3991/ijoe.v21i02.52909.
- [15] K. Sarode, R. Savdekar, and T. Chaudhari, "Texture Feature Analysis of an Image Using Gray Level Co-Occurrence Matrix," 2022. [Online]. Available: www.ijnrd.org
- [16] M. Yunus, "Feature Extraction : Gray Level Co-occurrence Matrix (GLCM) | by Muhammad Yunus | Medium." Accessed: Apr. 18, 2025. [Online]. Available: <https://yunusmuhammad007.medium.com/feature-extraction-gray-level-co-occurrence-matrix-glcm-10c45b6d46a1>
- [17] L. A. Andika, H. Pratiwi, and S. S. Handajani, "Klasifikasi Penyakit Pneumonia Menggunakan Metode Convolutional Neural Network Dengan Optimasi Adaptive Momentum," 2019.
- [18] F. Nurona Cahya, N. Hardi, D. Riana, and S. Hadiani, "SISTEMASI: Jurnal Sistem Informasi Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN)," Sep. 2021. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [19] P. A. K. K. Diallo and Y. Ju, "Accurate Detection COVID-19 Using K-EfficientNet Deep Learning Image Classifier dan K-COVID Chest X-Ray Images Dataset," *IEEE 6th International Conference on Computer and Communications*, pp. 1527–1531, 2020.
- [20] G. Celik, "Detection of Covid-19 and Other Pneumonia Cases from CT and X-ray Chest Images Using Deep Learning Based on Feature Reuse Residual Block and Depthwise Dilated Convolutions Neural Network," *Appl Soft Comput*, vol. 133, Jan. 2023, doi: 10.1016/j.asoc.2022.109906.
- [21] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," Mar. 2020.