

Perbandingan Klasifikasi Citra CT-Scan Kanker Paru-Paru Menggunakan Image Enhancement CLAHE Pada EfficientNet-B0

Dzaky Abdillah Salafy, Febi Yanto*, Surya Agustian, Fitri Insani

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

Email: ¹11950115039@students.uin-suska.ac.id, ^{2,*}febiyanto@uin-suska.ac.id, ³surya.agustian@uin-suska.ac.id,

⁴fitri.insani@uin-suska.ac.id,

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

Abstrak—Dalam beberapa tahun terakhir, peningkatan jumlah kasus kematian akibat kanker di dunia mengalami kenaikan yang cukup tinggi. Dari beberapa jenis kanker, kanker paru-paru adalah salah satu kasus tertinggi diantaranya. Kanker paru-paru umumnya diderita oleh laki-laki yang diakibatkan oleh beberapa hal seperti, asap rokok, polusi udara dalam jangka panjang, paparan terhadap senyawa yang bersifat karsinogenik, seperti radon, asbestos, arsenik, tar batubara, dan emisi bahan bakar diesel. Pertumbuhan sel kanker pada paru-paru dapat dideteksi menggunakan beberapa pemindaian citra salah satunya yaitu CT-Scan. Pada penelitian ini dilakukan klasifikasi organ paru-paru normal dan yang terjangkit oleh sel kanker. Proses pengklasifikasian menggunakan dua jenis data yaitu data asli dan data yang telah diproses dengan Contrast Limited Adaptive Histogram Equalization (CLAHE). Data akan terlebih dahulu dibagi dengan rasio 90:10 sebelum dilatih dengan menggunakan Convolutional Neural Network (CNN). Arsitektur CNN yang digunakan yaitu EfficientNet-B0 dibantu dengan optimizer dan learning rate yang berbeda. Setelah diuji model dievaluasi menggunakan confusion matrix untuk membandingkan performa antara penggunaan data asli dan data CLAHE. Penggunaan data CLAHE dengan clip limit 2 dan tile grid 8x8 mampu meraih hasil evaluasi lebih tinggi dibandingkan data asli yakni f1-score 84,45%, dan accuracy 84,71% pada pembagian 90:10, optimizer Adam, dan learning rate 10^{-1} . Hasil penelitian menunjukkan bahwa penggunaan image enhancement CLAHE dengan kombinasi yang tepat antara clip limit dan tile grid dapat memengaruhi performa model dalam mengklasifikasikan data citra.

Kata Kunci: CLAHE; CT-Scan; EfficientNet-B0; Kanker; Klasifikasi; Paru-paru

Abstract—In recent years, there has been a significant increase in the global cancer related mortality rate. Among various cancer types, lung cancer has emerged as one of the highest incidence cases. Lung cancer predominantly affects males and is attributed to several factors, including exposure to cigarette smoke, long-term air pollution, and exposure to carcinogenic compounds such as radon, asbestos, arsenic, coal tar, and diesel fuel emissions. The growth of cancerous cells in the lungs can be detected using various imaging techniques, with CT-Scan being one of them. This research focuses on the classification of normal lung organs and those affected by cancerous cells. The classification process employs two types of data: original data and data processed with Contrast Limited Adaptive Histogram Equalization (CLAHE). The data is initially divided with 90:10 ratios before being trained using a Convolutional Neural Network (CNN). The CNN architecture used is EfficientNet-B0, with the assistance of different optimizers and learning rates. After testing, the model's performance is evaluated using a confusion matrix to compare the results between the use of original data and CLAHE-processed data. The use of CLAHE processed data yields higher evaluation metrics compared to the original data, achieving a precision of 87.9%, recall of 85.6%, F1-score of 85.11%, and accuracy of 85.29% in the 90:10 data split, with the Adam optimizer and a learning rate of 10^{-1} . The research results reveal that the utilization of image enhancement, specifically Contrast Limited Adaptive Histogram Equalization (CLAHE), with an appropriate combination of clip limit and tile grid, can impact the model's performance in classifying image data.

Keywords: CLAHE; CT Scan; EfficientNet-B0; Cancer; Classification; Lungs

1. PENDAHULUAN

Menurut laporan World Health Organization (WHO) dari studi Global Burden of Cancer (Globocan) jumlah kematian akibat kanker di Indonesia mencapai 234.511 orang pada tahun 2020 [1]. Jumlah kasus baru pada kanker paru-paru menempati urutan ke-3 (8,8%), setelah kanker payudara (16,6%), dan kanker serviks (9,2%). Kanker paru merupakan jenis kanker yang paling banyak diderita oleh laki-laki (14,1%) dan menjadi salah satu penyebab kematian karena kanker dengan perkiraan 1,8 juta kasus kematian (18%), diikuti kolorektal (9,4%), liver (8,3%), lambung (7,7%), dan kanker payudara (6,9%) [2]. Dilansir dari Verywell Health, berdasarkan studi 2017 di BMC Public Health, selain asap rokok, kanker paru pada laki-laki ada kaitannya dengan paparan terhadap senyawa yang bersifat karsinogenik, seperti radon, asbestos, arsenik, tar batubara, dan emisi bahan bakar diesel [3].

Kasus kanker paru-paru meningkat setiap tahunnya, hal ini dibuktikan dengan statistik pertumbuhan kanker paru-paru dari 1.4 per 1000 penduduk di tahun 2013 menjadi 1,79 per 1000 penduduk pada tahun 2018 [4]. Umumnya, kanker paru-paru ditandai dengan gejala batuk, nyeri di dada, peradangan, dan masalah pernapasan lainnya. Jika gejala yang diderita semakin parah, dokter akan memeriksa organ paru-paru untuk dicek adanya sel abnormal atau tidak. Prosedur medis yang dilakukan untuk mendiagnosis adanya pertumbuhan sel abnormal pada paru-paru, ialah tes pencitraan, sitologi dahak, dan sampel jaringan (*biopsy*) [5]. Pemindaian tomografi menggunakan komputer atau CT-Scan merupakan salah satu tes pencitraan yang dapat digunakan untuk mendiagnosis kanker paru-paru.

Pada citra medis paru-paru yang terkena kanker, dapat ditemukan bercak putih mengumpul atau bisa disebut juga sebagai *lung nodule* [6]. Pendeteksian *nodule* secara manual membutuhkan waktu sekitar 15-20 menit untuk setiap citra medis. Hal ini tentu saja cukup melelahkan apabila mendeteksi citra medis dalam jumlah yang besar. Berdasarkan hal tersebut, banyak penelitian yang dilakukan untuk mendeteksi citra medis pada kasus kanker atau tumor pada organ tubuh manusia secara otomatis.

Penelitian terkait pengolahan citra medis telah dilakukan diantaranya, klasifikasi kanker kulit menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) oleh Muhammad Faruk, dkk[7], dan ekstraksi fitur citra MRI otak menggunakan Data Wavelet Transform (DWT) untuk klasifikasi penyakit tumor otak [8].

Seiring berkembangnya teknologi, dilakukan pengembangan terhadap *Machine Learning*, yaitu *Deep Learning*. Penggunaan *Deep Learning* dalam klasifikasi citra diklaim memiliki akurasi yang lebih baik dibandingkan dengan *Machine Learning*. Hal ini dibuktikan dengan penelitian oleh Mohammad Farid Naufal [9] yang membandingkan performa antara algoritma *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Convolutional Neural Network* (CNN). Hasil dari penelitian tersebut menyatakan bahwa, CNN yang merupakan salah satu algoritma dari *Deep Learning* memiliki hasil akurasi yang lebih baik dibandingkan KNN dan SVM dengan memperoleh *accuracy* 0,942.

Berdasarkan perbandingan terhadap beberapa algoritma untuk klasifikasi gambar yang telah dilakukan, maka penelitian ini akan menggunakan CNN dikarenakan lebih unggul dalam memperoleh nilai *accuracy*. CNN adalah salah satu algoritma *deep learning* yang merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. Struktur pada algoritma CNN, yaitu *convolutional layer*, *pooling layer*, dan terakhir *fully connected layer*. Pada saat ini, CNN umum digunakan pada pengolahan citra untuk klasifikasi, segmentasi, dan deteksi [10]. Algoritma ini telah digunakan pada beberapa penelitian terkait pengenalan citra seperti pada pengklasifikasian daging [11][12][13].

Salah satu arsitektur yang ada pada CNN adalah EfficientNet. Pada penelitian oleh Mingxing Tan dan Quoc V. Le [14] menjelaskan bahwa EfficientNet memiliki keunggulan dibandingkan dengan arsitektur seperti, ResNet dan DenseNet. Hal ini dikarenakan EfficientNet merupakan arsitektur yang sederhana dan memiliki jumlah parameter yang lebih sedikit namun, dapat memperoleh akurasi yang lebih baik. Di antara model B0-B7, EfficientNet-B0 merupakan model yang memiliki jumlah parameter paling kecil yaitu 5,3 juta sehingga waktu komputasi lebih cepat. Banyak penelitian yang telah menggunakan EfficientNet-B0 seperti, pendeteksian penyakit pada daun tumbuhan oleh Sk Mahmudul Hassan, dkk [15]. Hasil dari penelitian ini menunjukkan bahwa EfficientNet-B0 mengungguli arsitektur lain dengan memperoleh akurasi lebih tinggi yaitu 99,56%. Kemudian penelitian oleh Wisnu Gilang Pamungkas[18], menggunakan EfficientNet-B0 dalam mengklasifikasi penyakit pada jagung yang memperoleh akurasi lebih tinggi dibandingkan ResNet yakni 94%.

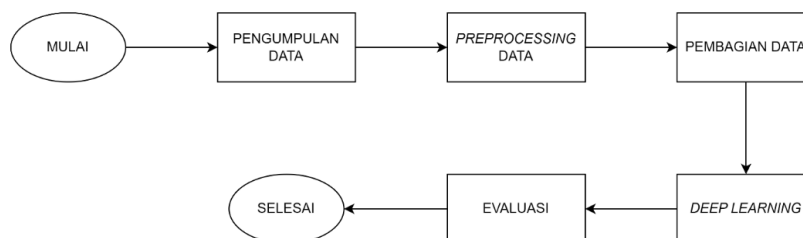
Pada tahun 2022, dilakukan penelitian terhadap pengaruh *image enhancement* dalam peningkatan akurasi arsitektur EfficientNet untuk klasifikasi citra CT-Scan COVID-19. Metode *image enhancement* atau *transform* yang digunakan pada penelitian ini antara lain, *Laplace Transform*, *Wavelet Transforms*, *Adaptive Gamma Correction*, dan *Contrast Limited Adaptive Histogram Equalization* (CLAHE). Hasil dari penelitian tersebut menyatakan bahwa *image enhancement* berpengaruh pada peningkatan akurasi EfficientNet, dan CLAHE merupakan metode *image enhancement* yang mendapatkan akurasi tertinggi yaitu 94.56% [17]. Didukung juga dengan beberapa penelitian lainnya seperti identifikasi penyakit pada citra *x-ray* pada paru-paru[18][19].

Oleh karena itu, penelitian ini menggunakan algoritma CNN dengan arsitektur EfficientNet-B0 untuk klasifikasi kanker paru-paru. Pada tahap *preprocessing* akan dilakukan *resize*, dan proses *image enhancement* untuk meningkatkan kualitas citra agar mendapatkan akurasi yang lebih akurat [10]. *Contrast Limited Adaptive Histogram Equalization* (CLAHE) adalah metode dari *image enhancement* yang akan digunakan, karena dapat meningkatkan kontras pada citra [17]. Penelitian ini juga akan menggunakan *Hyperparameter Optimization* jenis Adam, *Stochastic Gradient Descent* (SGD), dan Adagrad untuk pengoptimalan kinerja CNN. Hal ini dikarenakan performa dari CNN dapat meningkat dan memberikan hasil yang lebih baik jika menggunakan *hyperparameter* yang tepat [20].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

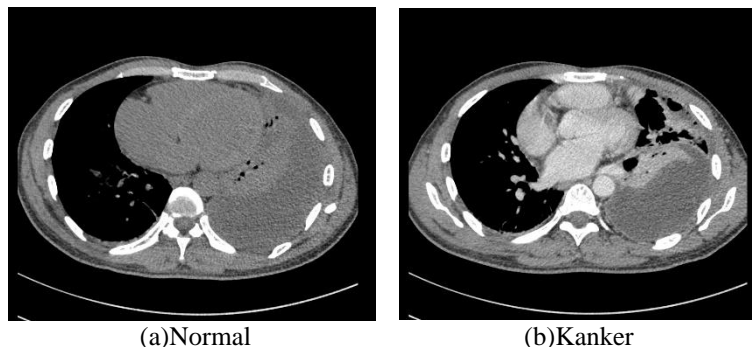
Alur dari penelitian ini ditunjukkan pada gambar 1. Tahapan yang dilakukan diantaranya ialah pengumpulan data, *preprocessing* data, *deep learning*, dan evaluasi. Tahapan pertama yaitu melakukan pengumpulan data yang nantinya akan digunakan untuk melatih model *deep learning*. Selanjutnya dilakukan *preprocessing* seperti, *resize* dan *image enhancement* agar data sesuai dengan dengan *input* pada *deep learning*. Kemudian data yang telah diproses pada dua langkah sebelumnya akan dilatih menggunakan *deep learning* dengan arsitektur EfficientNet-B0 yang dibantu dengan *hyperparameter optimization* untuk meningkatkan performa pelatihan. Terakhir, data yang telah dilatih akan dievaluasi menggunakan *confusion matrix*.



Gambar 1. Workflow Penelitian

2.2 Pengumpulan Data

Tahap pertama yang dilakukan pada penelitian ini adalah mengumpulkan data citra CT-Scan paru-paru. Data sekunder citra paru-paru pasien RSUD Arifin Achmad Provinsi Riau. Kemudian data diseleksi agar memberikan variasi pada citra yang dilatih. Setelah diseleksi, data yang digunakan berjumlah 1537 dan terbagi menjadi dua kelas, yaitu normal dan kanker. Citra normal sebanyak 749 gambar dan citra kanker sebanyak 788 gambar.



Gambar 2. Perbandingan citra normal dan kanker

2.3 Preprocessing Data

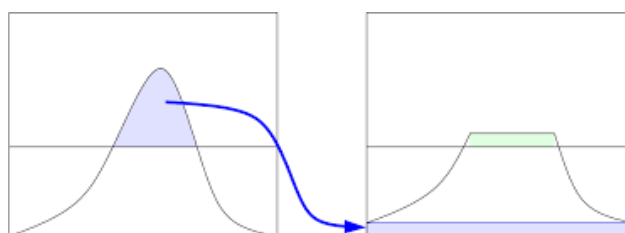
Tahap selanjutnya adalah melakukan *preprocessing* terhadap data citra yang telah dikumpulkan. Data yang telah dikumpulkan akan di-*resize* dan ditingkatkan kontrasnya menggunakan proses *image enhancement* dengan metode *Contrast Limited Adaptive Histogram Equalization* (CLAHE).

a. Resize

Data CT-Scan paru-paru akan di *resize* menjadi 224×224 *pixel*. Hal ini dilakukan karena model EfficientNet-B0 membutuhkan *input* sebesar 224×224 *pixel* sehingga data dapat dilatih dengan model tersebut.

b. Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE dilakukan untuk meningkatkan *contrast* pada citra yang nantinya akan diproses pada *Deep Learning*. Metode ini akan memberikan nilai batas atau *clip limit* untuk membatasi amplifikasi pada histogram sehingga dapat mengurangi masalah derajat keabuan pada citra. Histogram asli setiap blok dipotong dari atas jika melebihi batas yang telah ditentukan atau *clip limit*. Kemudian, bagian histogram yang dipotong, akan didistribusikan secara merata pada seluruh rentang *pixel* histogram.



Gambar 3. Ilustrasi CLAHE

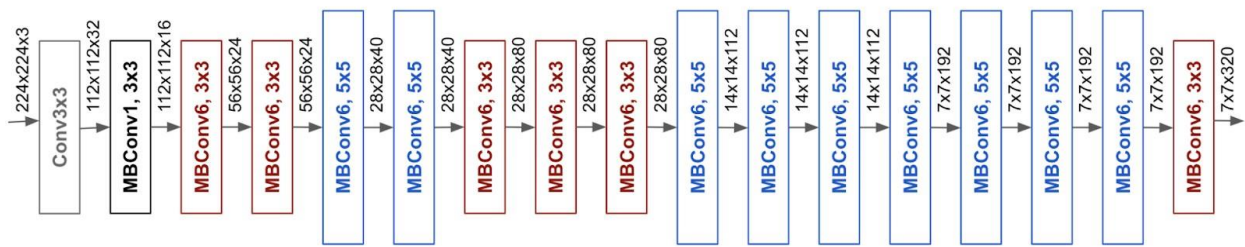
Gambar 3 merupakan ilustrasi dari *Contrast Limited Adaptive Histogram Equalization* (CLAHE) yang akan diterapkan melalui *library* OpenCV dengan melibatkan dua parameter, yaitu *clip limit* dan *tile grid size*. Proses CLAHE dimulai dengan mengatur ukuran wilayah dan batas pemotongan. Selanjutnya, histogram dibentuk untuk setiap wilayah sesuai dengan *tile grid size* yang ditentukan, kemudian histogram tersebut dipotong menggunakan batas pemotongan atau *clip limit* yang telah ditetapkan, dan nilai-nilai yang melebihi batas pemotongan didistribusikan kembali ke bagian lain dari histogram [21]. Setelah itu, histogram yang baru dihasilkan diterapkan pada gambar, dan dilakukan pengisian piksel di wilayah tetangga.

2.4 Pembagian Data

Data akan dibagi menjadi dua kategori berdasarkan jenisnya yaitu, data asli dan data CLAHE. Selanjutnya kedua jenis data tersebut, masing-masing akan dibagi menjadi data latih dan data uji. Data citra akan dibagi terlebih dahulu antara data latih dan data uji dengan rasio 90:10. Kemudian, sebagian data latih yang telah dibagi sebelumnya akan digunakan untuk data validasi dengan rasio yaitu 90:10.

2.5 Deep Learning

Data citra yang sudah diproses pada tahap sebelumnya, selanjutnya akan diklasifikasi menggunakan *Deep Learning*. Penelitian ini menggunakan *Deep Learning* algoritma *Convolutional Neural Network* (CNN) arsitektur EfficientNet-B0. Berikut adalah gambar arsitektur dari EfficientNet-B0:



Gambar 4. Arsitektur EfficientNet-B0

Gambar 4 merupakan arsitektur dari EfficientNet-B0. Berikut penjelasan tiap *layer* atau lapisan yang ada pada arsitektur tersebut:

- Convolutional Layer*: Lapisan yang berisi sejumlah filter yang digunakan untuk mendeteksi input berupa gambar atau citra dengan melakukan proses konvolusi.
- Intermediate Layer*: Terdiri dari beberapa lapisan yang digunakan untuk mengesktrasi fitur-fitur kompleks yang ada pada gambar.
- Pooling Layer*: Lapisan yang digunakan untuk menyederhanakan dimensi data kemudian merangkum informasi penting dari peta fitur yang dibuat pada lapisan sebelumnya.
- Fully Connected Layer*: Memiliki hubungan neuron dengan lapisan sebelumnya yang memuat informasi penting untuk dilakukan klasifikasi dengan cara mengekstaksi fitur-fitur sesuai kategori atau kelas yang diprediksi.
- Output Layer*: Merupakan lapisan terakhir yang berisi hasil prediksi akhir dari lapisan sebelumnya.

Tabel 1. Hyperparameter Optimization

<i>Optimizer</i>	<i>Learning Rate</i>
Adam	0,1
SGD	0,01
Adagrad	0,001
	0,0001

Tabel 1 menggambarkan bahwa pelatihan model dengan EfficientNet-B0 ini akan menggunakan *Hyperparameter Optimization* jenis Adagrad, *Stochastic Gradient Descent* (SGD), dan Adam untuk pengoptimalan kinerja CNN. Hal ini dikarenakan performa dari CNN dapat meningkat dan memberikan hasil yang lebih baik jika menggunakan *hyperparameter* yang tepat [20].

2.6 Evaluasi

Model akan diuji dengan data *test* yang telah dipisahkan sebelumnya. Data *test* dibagi menjadi dua, sesuai dengan kategori yang juga digunakan pada data *train* atau latih sebelumnya, yaitu kanker dan normal. Jumlah data *test* kanker yaitu 87 sedangkan untuk normal berjumlah 83. Kemudian akan dievaluasi dengan *confusion matrix*. *Confusion matrix* digunakan untuk membandingkan berapa banyak jumlah data yang sesuai antara *input* dengan prediksi. Ada beberapa istilah yang digunakan pada *confusion matrix*, yaitu:

- True Positive* (TP): Jumlah data positif yang diklasifikasikan benar oleh model.
- True Negative* (TN): Jumlah data negatif diklasifikasikan benar oleh model.
- False Positive* (FP): Jumlah data positif yang diklasifikasikan salah oleh model.
- False Negative* (FN): Jumlah data negatif yang diklasifikasikan salah oleh model.

Berikut nama-nama metrik dan rumusnya yang terdapat pada *confusion matrix*:

- Precision*: Mengukur sejauh mana model dapat mengidentifikasi *input* dengan benar sesuai kelasnya.

$$Precision = \frac{TP}{TP+FN} \times 100\% \tag{1}$$

- Recall*: Menghitung berapa persen ketepatan model dalam memprediksi kelas positif yang sebenarnya.

$$Recall = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

- F1-Score*: Mengukur keseimbangan antara nilai *precision* dan *recall* dengan cara menyatukan keduanya menjadi satu nilai

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{3}$$

- Accuracy*: Menghitung ketepatan model dalam memprediksi benar semua kelas yang diuji.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Model Optimal

Proses ini dilakukan untuk menentukan kombinasi yang tepat pada nilai *pooling* dan *dense layers* yang akan digunakan pada model EfficientNet-B0 untuk proses pelatihan data. Penentuan model ini dilakukan dengan cara mencari nilai *f1-score* dan *accuracy* terbaik terhadap validasi pada data asli.

Tabel 2. Hasil Pemilihan *Dense Layers*

No	Dense Layers	Learning Rate	F1 Score	Akurasi
1	128,64,32,1	0,1	41,95	42,12
2	64,32,16,1	0,1	38,65	38,87
3	32,16,8,1	0,1	48,61	48,97
4	32,8,8,1	0,1	51,43	52,10
5	16,8,8,1	0,1	54,15	54,57
6	16,8,4,1	0,1	42,33	42,45
7	32,8,1	0,1	46,05	46,25
8	16,8,1	0,1	47,72	47,77

Tabel 2 menampilkan hasil berbagai skenario untuk menetapkan terlebih dahulu nilai dari *dense layers* yang akan digunakan seterusnya pada proses pelatihan. Dari tabel tersebut didapatkan kombinasi menggunakan *dense layer* 16, 8, 8, 1 memperoleh hasil terbaik dengan perolehan *f1-score* sebesar 54,15% dan *accuracy* 56,2%. Kemudian dari *dense layers* tersebut akan ditentukan nilai *pooling* yang tepat.

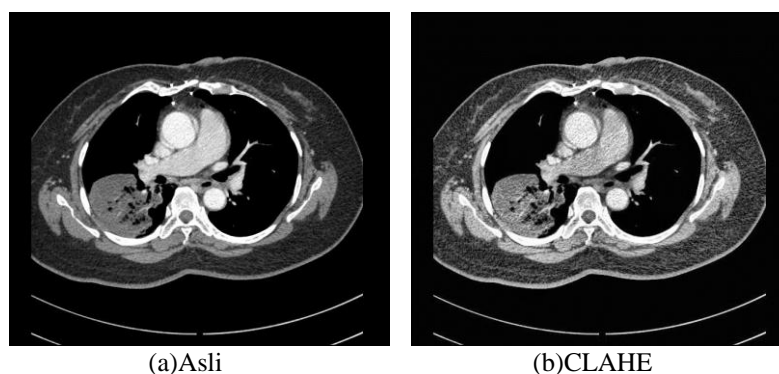
Tabel 3. Hasil Pemilihan *Pooling*

No	Pooling	Optimizer	Dense Layers	Learning Rate	F1 Score	Akurasi
1	Avg	Adam	16,8,8,1	0,1	45,14	45,99
2	Max	Adam	16,8,8,1	0,1	54,15	56,20
3	Avg	SGD	16,8,8,1	0,1	48	48,18
4	Max	SGD	16,8,8,1	0,1	49,59	50,36
5	Avg	Adagrad	16,8,8,1	0,1	46,81	47,45
6	Max	Adagrad	16,8,8,1	0,1	44,50	45,26

Tabel 3 menunjukkan perolehan terbaik dengan menggunakan *pooling* Max yang dikombinasikan dengan *optimizer* Adam dan *dense layers* 16, 8, 8, 1 memperoleh hasil terbaik dengan *f1-score* sebesar 54,15% dan *accuracy* 56,2% terhadap data validasi.

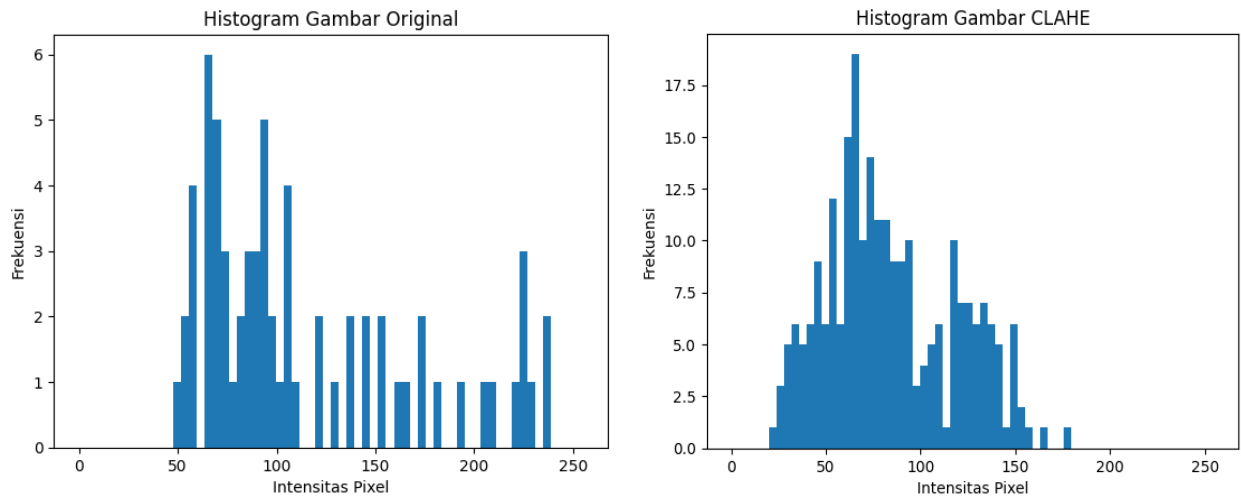
3.2 Penerapan CLAHE pada Citra

Sebagian percobaan yang dilakukan menggunakan data latih yang telah diberikan CLAHE. Penerapan CLAHE pada citra bertujuan untuk meningkatkan kontras. Hal ini dilakukan karena citra yang digunakan untuk pelatihan adalah citra CT-Scan yang skala warnanya *grayscale* atau hanya memiliki tingkat keabuan. Peningkatan kontras diharapkan dapat membantu model dalam mendapatkan informasi yang lebih spesifik dalam proses pengenalan citra. Pada proses ini, digunakan *library* OpenCV untuk memanfaatkan fitur CLAHE yang sudah disediakan agar dapat langsung digunakan dengan mudah. Ada beberapa faktor yang harus diperhatikan dalam penerapan CLAHE pada citra seperti, *clip limit*, dan *tile grid*. Namun, citra yang sebelumnya memiliki 3 *channel* warna yaitu *blue*, *green*, *red* (BGR) akan diubah terlebih dahulu menjadi *grayscale* sebelum melakukan penerapan CLAHE. *Clip limit* digunakan untuk mengatur batas kontras sedangkan *tile grid* untuk menetapkan ukuran *grid* pada pemerataan histogram. Pemberian CLAHE pada penelitian ini menggunakan *clip limit* sebesar 2 dan *tile grid* berukuran 8x8. Berikut perbandingan antara citra asli dan citra yang telah ditingkatkan kontrasnya dengan CLAHE.



Gambar 5. Perbandingan citra asli dan CLAHE

Dari gambar 5 di atas terlihat perbedaan kontras antara gambar asli dan yang telah diproses dengan CLAHE, jika dilihat nilai histogramnya akan menghasilkan grafik seperti ini:



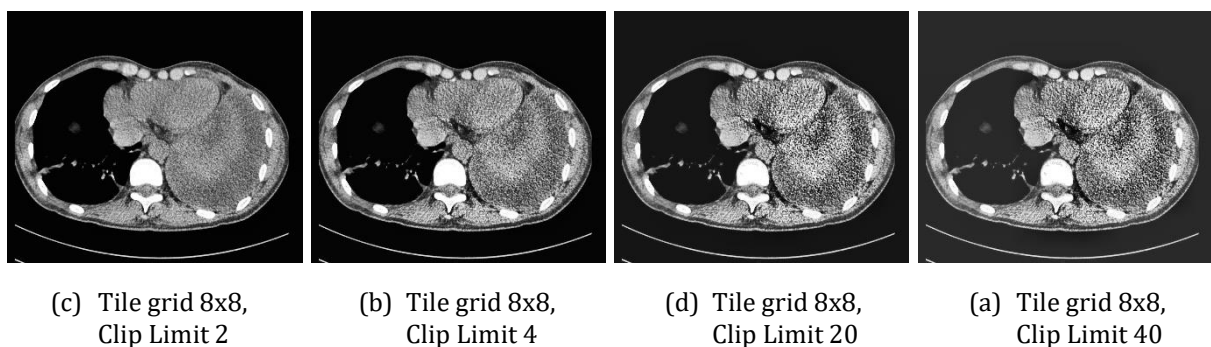
Gambar 6. Perbandingan histogram citra asli dan CLAHE

Dari grafik pada gambar 6, dapat dilihat bahwa citra yang telah ditingkatkan kontrasnya dengan CLAHE memiliki intensitas piksel yang lebih merata dibandingkan dengan citra asli. Hal ini dibuktikan dengan meningkatnya frekuensi pada beberapa intensitas piksel. Ada beberapa nilai yang akan dicoba pada penerapan CLAHE untuk citra CT-Scan paru-paru kanker dan normal di penelitian ini, di antaranya:

Tabel 5. Nilai Tiap Parameter CLAHE yang diterapkan

No	Tile Grid	Clip Limit
1	8x8	2
2	8x8	4
3	8x8	20
4	8x8	40
5	16x16	2
6	16x16	4
7	16x16	20
8	16x16	40
9	32x32	2
10	32x32	4
11	32x32	20
12	32x32	40

Setelah nilai pada tabel 2 di atas diterapkan pada citra CT-Scan paru-paru, maka akan menghasilkan citra yang lebih kontras dibandingkan citra asli. Berikut ini beberapa citra hasil dari penerapan CLAHE.



Gambar 7. Citra CLAHE

3.3 Proses Training

Citra yang telah melalui proses pembagian data, akan dilatih menggunakan model EfficientNet-B0 yang disediakan oleh library Tensorflow. Dalam proses pelatihan ini, model EfficientNet-B0 akan ditambahkan *hyperparameter optimization* untuk meningkatkan performa model pada saat mengenali citra. Ada tiga *optimizer* yang digunakan pada proses ini yaitu,

Adam, SGD, dan Adagrad. Setiap *optimizer* akan dilakukan percobaan dengan empat *learning rate* yang berbeda dari 10^{-1} hingga 10^{-4} . Ukuran *batch* yang digunakan ialah 64 untuk setiap kali percobaan. Pada setiap pelatihan akan dilakukan sebanyak 50 *epoch* dengan ditambah *early stopping* dan *checkpoint*. *Early stopping* digunakan untuk menghentikan proses pelatihan ketika akurasi validasi tidak meningkat selama 10 *epoch*. Sedangkan *checkpoint* digunakan untuk menyimpan akurasi validasi terbaik selama proses pelatihan berlangsung. Berikut hasil pelatihan data asli dan CLAHE dengan *clip limit* 2 dan *tile grid* 8x8 pada pembagian 90:10. Setelah proses pelatihan selesai, data akan dievaluasi menggunakan *confusion matrix* untuk mendapatkan nilai *precision*, *recall*, *f1-score*, dan *accuracy* pada model.

Tabel 6. Hasil Train Data Asli dan Data CLAHE

No	Jenis Data	Optimizer	Learning Rate	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	Asli	Adam	0,1	0,3819	0,7748	0,3807	0,8102
2	Asli	Adam	0,01	0,3742	0,7732	0,4449	0,7956
3	Asli	Adam	0,001	0,3853	0,7659	0,4461	0,7737
4	Asli	Adam	0,0001	0,3766	0,7675	0,3935	0,8029
5	Asli	SGD	0,1	0,4033	0,7675	0,3792	0,8102
6	Asli	SGD	0,01	0,4183	0,7610	0,4312	0,7810
7	Asli	SGD	0,001	0,4492	0,7528	0,4570	0,7956
8	Asli	SGD	0,0001	0,4592	0,7512	0,4696	0,7591
9	Asli	Adagrad	0,1	0,4624	0,7480	0,4635	0,7737
10	Asli	Adagrad	0,01	0,4984	0,7374	0,4795	0,7299
11	Asli	Adagrad	0,001	0,5140	0,7154	0,5583	0,6496
12	Asli	Adagrad	0,0001	0,5366	0,7293	0,4835	0,7664
13	CLAHE	Adam	0,1	0,3777	0,7659	0,3836	0,8321
14	CLAHE	Adam	0,01	0,3682	0,7764	0,4111	0,7591
15	CLAHE	Adam	0,001	0,3725	0,7683	0,4307	0,7664
16	CLAHE	Adam	0,0001	0,3853	0,7667	0,4654	0,7664
17	CLAHE	SGD	0,1	0,4294	0,7659	0,4085	0,8248
18	CLAHE	SGD	0,01	0,4219	0,7585	0,4339	0,7956
19	CLAHE	SGD	0,001	0,3975	0,7675	0,4148	0,7664
20	CLAHE	SGD	0,0001	0,3703	0,774	0,4408	0,708
21	CLAHE	Adagrad	0,1	0,4739	0,7439	0,4845	0,708
22	CLAHE	Adagrad	0,01	0,4915	0,7293	0,4699	0,7518
23	CLAHE	Adagrad	0,001	0,5689	0,7089	0,5827	0,6788
24	CLAHE	Adagrad	0,0001	0,5192	0,7179	0,4784	0,7883

Tabel 6 mencerminkan hasil yang diperoleh selama proses pelatihan, di mana pelatihan dilakukan terhadap data latih dan data validasi dengan berbagai skenario untuk mengamati variasi hasil pelatihan model. Tabel tersebut menyajikan informasi terkait jenis data, *optimizer*, *learning rate*, *training loss*, *training accuracy*, *validation loss* dan *validation accuracy*. *Training loss* dan *validation loss* metrik yang digunakan untuk mengukur seberapa baik model machine learning atau deep learning melakukan prediksi terhadap data yang tidak dilihat selama pelatihan. Sementara itu, *training accuracy* dan *validation accuracy* mengukur kemampuan model dalam mengklasifikasikan data secara tepat. Berikut adalah beberapa visualisasi grafik yang merepresentasikan proses pelatihan model.

3.4 Perbandingan Hasil Data Asli dan CLAHE

Setelah menguji dua jenis citra yaitu, citra asli dan CLAHE (Contrast Limited Adaptive Histogram Equalization) pada pembagian data 90:10 maka akan dilakukan perbandingan terhadap keduanya. Perbandingan dilakukan untuk mengukur pengaruh yang diberikan oleh image enhancement terhadap pelatihan sebuah model dalam mengklasifikasi citra. Perbandingan antara citra yang telah diolah dengan teknik CLAHE dan citra asli merupakan aspek yang signifikan dalam pengolahan citra dan analisis visual. Tujuan utamanya adalah untuk mengukur dan mengevaluasi dampak penggunaan CLAHE dalam meningkatkan kualitas citra. Salah satu aspek yang paling terlihat adalah peningkatan kontras dan kejelasan citra. CLAHE dapat efektif mengatasi masalah kontras yang rendah atau ketidakseimbangan dalam citra asli, sehingga memberikan gambar yang lebih tajam dan lebih jelas. Namun, perlu juga diperhatikan bahwa penggunaan CLAHE tidak selalu memberikan hasil yang lebih baik dalam setiap situasi. Ada kemungkinan bahwa dalam beberapa kasus, pengolahan dengan CLAHE dapat menghasilkan peningkatan noise dalam citra, terutama jika parameter yang digunakan tidak diatur dengan baik. Selain itu, perbandingan ini juga membantu dalam menentukan konteks aplikasi yang paling cocok untuk penggunaan CLAHE. Misalnya dalam citra medis, teknik ini mungkin sangat berguna untuk meningkatkan visualisasi detail anatomi. Dalam keseluruhan, perbandingan antara citra CLAHE dan asli adalah langkah penting dalam pengambilan keputusan dalam pengolahan citra dan membantu memilih metode yang paling sesuai untuk tujuan tertentu.

Tabel 7. Pengujian Data Asli

No	Optimizer	Learning Rate	F1-Score	Accuracy
1	Adam	0,1	81,69	81,76
2	Adam	0,01	79,96	80
3	Adam	0,001	77,04	77,06
4	Adam	0,0001	70,28	71,76
5	SGD	0,1	78,23	78,24
6	SGD	0,01	80,58	80,59
7	SGD	0,001	79,99	80
8	SGD	0,0001	79,03	79,41
9	Adagrad	0,1	76,63	77,06
10	Adagrad	0,01	75,71	76,47
11	Adagrad	0,001	70,65	72,35
12	Adagrad	0,0001	80,29	80,59

Setelah dilakukan pengujian terhadap data asli yang ditunjukkan pada tabel 7, maka didapatkan hasil tertinggi oleh penggunaan *optimizer* Adam dengan *learning rate* 10^{-1} . Hasilnya yaitu, *f1-score* 81,69% dan *accuracy* 81,76%. Selanjutnya akan dilakukan pengujian menggunakan dataset CLAHE untuk mengetahui pengaruhnya dalam meningkatkan *f1-score* dan *accuracy*. Berikut ini hasil pengujian menggunakan beberapa kombinasi CLAHE dengan *optimizer* Adam dan *learning rate* 10^{-1} .

Tabel 8. Pengujian Data CLAHE

No	Tile Grid	Clip Limit	F1-Score	Accuracy
1	8x8	2	84,45	84,71
2	8x8	4	76,2	76,37
3	8x8	20	81,13	81,18
4	8x8	40	82,15	82,35
5	16x16	2	74,39	74,71
6	16x16	4	76,42	76,5
7	16x16	20	81,58	81,76
8	16x16	40	79,96	80
9	32x32	2	73,04	73,53
10	32x32	4	83,87	84,12
11	32x32	20	72,56	72,94
12	32x32	40	82,26	82,35

Tabel 8 menampilkan nilai *f1-score* dan *accuracy* saat menggunakan CLAHE. Nilai paling tinggi diraih oleh penggunaan *clip limit* 2 dan *tile grid* 8x8 dibantu dengan *optimizer* Adam dan *learning rate* 10^{-1} yang memperoleh *f1-score* 84,45%, dan *accuracy* 84,71%.

4. KESIMPULAN

Pada penelitian ini dilakukan pengklasifikasian citra kanker paru-paru normal dan kanker menggunakan CNN EfficientNet-B0. Data yang dilatih terbagi menjadi dua jenis yaitu citra asli dan citra CLAHE yang sama-sama mendapatkan pembagian data dengan rasio 90:10. Total data citra yang digunakan berjumlah 1537 dan terbagi menjadi 2 kategori yaitu normal sebanyak 749 gambar dan kanker sebanyak 788 gambar. Setiap model dilatih dengan *batch size* 64 dan *epoch* 50. Kemudian performa model yang telah dilatih, diukur menggunakan *confusion matrix* untuk mendapatkan nilai dari *precision*, *recall*, *f1-score*, dan *accuracy*. Dari percobaan yang telah dilakukan pada data asli didapatkan hasil uji tertinggi pada pembagian rasio 90:10 menggunakan *optimizer* Adam dengan *learning rate* 10^{-1} mendapatkan nilai *f1-score* 81,69%, dan *accuracy* 81,76%. Sementara itu, hasil uji tertinggi dengan pelatihan data CLAHE terdapat pada kombinasi *clip limit* 2 dan *tile grid* 8x8 menggunakan *optimizer* Adam dengan *learning rate* 10^{-1} mendapatkan nilai *f1-score* 84,45%, dan *accuracy* 84,71%. Oleh karena itu, dapat disimpulkan bahwa pembagian data dapat memengaruhi performa model dalam mengklasifikasi citra begitu juga dengan penggunaan CLAHE dalam meningkatkan citra *grayscale* khususnya citra CT-Scan untuk memberikan informasi kepada model saat dilatih. Disarankan untuk memperbanyak data citra yang akan dilatih ataupun menggunakan kombinasi *clip limit* dan *tile grid* yang lebih bervariasi, dan bisa juga menggunakan beberapa jenis *image enhancement* lainnya seperti, *Gamma Correction* atau *Wavelet Transform*.

REFERENCES

- [1] Advertorial, "Kanker Paru-paru: 7 Hal yang Perlu Anda Ketahui," www.pikiran-rakyat.com. Accessed: Dec. 08, 2022. [Online]. Available: <https://www.pikiran-rakyat.com/bandung-raya/pr-015756667/kanker-paru-7-hal-yang-perlu-anda-ketahui>

- [2] Kurniyanto, “Bagaimana Kanker Paru dapat Diketahui Lebih Awal Sebelum Stadium Lanjut?,” www.yankes.kemkes.go.id. Accessed: Dec. 08, 2022. [Online]. Available: [https://yankes.kemkes.go.id/view_artikel/1550/bagaimana-kanker-paru-dapat-diketahui-lebih-awal-sebelum-stadium-lanjut#:~:text=Di Indonesia%2C berdasarkan data Globocan,laki \(14%2C1%25\)](https://yankes.kemkes.go.id/view_artikel/1550/bagaimana-kanker-paru-dapat-diketahui-lebih-awal-sebelum-stadium-lanjut#:~:text=Di Indonesia%2C berdasarkan data Globocan,laki (14%2C1%25)).
- [3] L. Eldridge, “Lung Cancer in Men,” www.verywellhealth.com. Accessed: Dec. 08, 2022. [Online]. Available: <https://www.verywellhealth.com/lung-cancer-in-men-2249258>
- [4] N. Handayani, “Kanker dan Serba-Serbinya (Hari Kanker Sedunia 2022).” Accessed: Dec. 18, 2022. [Online]. Available: <https://tsprespira.jogjaprovo.go.id/kanker-dan-serba-serbinya-hari-kanker-sedunia-2022/>
- [5] R. Fadli, “Kanker Paru.” Accessed: Dec. 18, 2022. [Online]. Available: <https://www.halodoc.com/kesehatan/kanker-paru>
- [6] N. Jawas and I. W. B. Sentana, “Deteksi Lung Nodule Otomatis dari Citra Medis Computed Tomography (CT) Scan dengan Operasi Morfology dan SVM,” *Jurnal Sistem Dan Informatika*, vol. 12, no. 2, pp. 140–148, 2018.
- [7] M. Faruk and N. Nafi’iyah, “Klasifikasi Kanker Kulit Berdasarkan Fitur Tekstur, Fitur Warna Citra Menggunakan SVM dan KNN,” *Telematika*, vol. 13, no. 2, pp. 100–109, 2020.
- [8] L. W. Astuti, “Ekstrasi Fitur Citra MRI Otak Menggunakan Data Wavelet Transform (DWT) untuk Klasifikasi Penyakit Tumor Otak,” *Jurnal Ilmiah Informatika Global*, vol. 10, no. 2, pp. 80–86, 2019, doi: 10.36982/jig.v10i2.854.
- [9] M. Farid Naufal, “Perbandingan, Analisis Svm, Algoritma Untuk, dan CNN,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 2, pp. 311–318, 2021, doi: 10.25126/jtiik.202184553.
- [10] D. Gunawan and H. Setiawan, “Convolutional Neural Network dalam Citra Medis,” *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 2, pp. 376–390, 2022, doi: 10.24002/konstelasi.v2i2.5367.
- [11] D. L. Y. Ikhwanul, J. Jasril, S. Sanjaya, L. Handayani, and F. Yanto, “Klasifikasi Citra Daging Sapi dan Babi Menggunakan CNN Alexnet dan Augmentasi Data,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, Jul. 2023, doi: 10.47065/josh.v4i4.3702.
- [12] S. Lasniari, J. Jasril, S. Sanjaya, F. Yanto, and M. Affandes, “Klasifikasi Citra Daging Babi dan Daging Sapi Menggunakan Deep Learning Arsitektur ResNet-50 dengan Augmentasi Citra,” *Jurnal Sistem Komputer Dan Informatika (JSON)*, vol. 3, no. 4, pp. 450–457, 2022.
- [13] M. F. Martias, J. Jasril, S. Sanjaya, L. Handayani, and F. Yanto, “Klasifikasi Citra Daging Sapi dan Daging Babi Menggunakan CNN Arsitektur EfficientNet-B6 dan Augmentasi Data,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, p. 642, Jun. 2023, doi: 10.30865/json.v4i4.6195.
- [14] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
- [15] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, “Identification of plant-leaf diseases using CNN and transfer-learning approach,” *Electronics (Basel)*, vol. 10, no. 12, p. 1388, 2021.
- [16] W. G. Pamungkas, M. I. P. Wardhana, Z. Sari, and Y. Azhar, “Leaf Image Identification: CNN with EfficientNet-B0 and ResNet-50 Used to Classified Corn Disease,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 326–333, Mar. 2023, doi: 10.29207/resti.v7i2.4736.
- [17] A. Shamila Ebenezer, S. Deepa Kanmani, M. Sivakumar, and S. Jeba Priya, “Effect of image transformation on EfficientNet model for COVID-19 CT image classification,” *Mater Today Proc*, vol. 51, pp. 2512–2519, 2022, doi: 10.1016/j.matpr.2021.12.121.
- [18] Dodi Andre Putra, J. Na’am, and Yuhandri, “Identifikasi Objek pada Citra Thorax X-Ray Pasien COVID-19 dengan Metode Contrast Limited Adaptive Histogram Equalization (CLAHE),” *Jurnal Informasi dan Teknologi*, vol. 4, pp. 33–38, 2022, doi: 10.37034/jidt.v4i1.184.
- [19] R. Moch Diar, R. Y. Fu’adah, and K. Usman, “Klasifikasi Penyakit Paru-Paru Berbasis Pengolahan Citra X Ray Menggunakan Convolutional Neural Network (Classification Of The Lung Diseases Based On X Ray Image Processing Using Convolutional Neural Network),” *e-Proceeding of Engineering*, vol. 9, no. 2, pp. 476–484, 2022.
- [20] T. Badriyah, D. B. Santoso, I. Syarif, and D. R. Syarif, “Improving stroke diagnosis accuracy using hyperparameter optimized deep learning,” *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 3, pp. 256–272, 2019, doi: 10.26555/ijain.v5i3.427.
- [21] F. M. Hana and I. D. Maulida, “Analysis of contrast limited adaptive histogram equalization (CLAHE) parameters on finger knuckle print identification,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1764/1/012049.