

joins

JOURNAL OF INFORMATION SYSTEM

e-ISSN :2528-0236

Vol.	No.
07	02
November 2022	



**UNIVERSITAS
DIAN NUSWANTORO**



9 772528 023007

Table of Contents

Articles

Evaluasi Kualitas Sistem Informasi Akademik Dengan Webqual dan IPA DOI : 10.33633/joins.v7i2.6187 Yuwan Jumaryadi, Misbahul Fajri, Bagus Priambodo	PDF 99-106
Klasifikasi Penentuan skema Uji Sertifikasi di LSP UDINUS bagi mahasiswa Progrdi Sistem Informasi UDINUS dengan Algoritma Decision Tree (C4.5) DOI : 10.33633/joins.v7i2.6429 Agus Winarno, Budi Warsito, Adi Wibowo, Junta Zeniarja	PDF 107-114
Penerapan Arsitektur Enterprise pada Pelayanan Pendaftaran Anggota pada Fitnation Premiere Gym DOI : 10.33633/joins.v7i2.6396 Indra Gamayanto, Galih Mentari Pangesti, Ramadhan Rakhmat Sani	PDF 115-123
Analisis Quality In Use Model ISO/IEC 25010 Pada Penggunaan Aplikasi TikTok DOI : 10.33633/joins.v7i2.6491 Masfufahatul Umroh, Syafrina Dyah Kusuma Wardani, Aulia Cahya Rani, Achmad Maulana Riski, M. Dwi Cahya Bimantara, Dwi Rolliawati	PDF 124-130
Penerapan Zachman Framework dalam Merancang Customer Relationship Management pada Bank Central Asia Indonesia DOI : 10.33633/joins.v7i2.6308 R. Budiraharjo, Prambudhi Wibowo Pandji, Boyke Agung Nugraha, Bahy Tsany Rizqullah4	PDF 131-142
Pengujian Kebergunaan Aplikasi Computer Assisted Test (CAT) Badan Kepegawaian Negara (BKN) Menggunakan System Usability Scale (SUS) DOI : 10.33633/joins.v7i2.6647 Teddy Cahyo Munanto, Sukma Suryaningsih	PDF 143-154
Penerapan Metode Kompresi Huffman untuk Meningkatkan Jumlah Data Informasi pada Kode Quick Response DOI : 10.33633/joins.v7i2.6789 Hidayat Hidayat, Muchammad Yaasiin Asshobry	PDF 155-165
Klasifikasi Malaria Menggunakan Metode Image Processing Dari Sel Darah Merah Dengan Algoritma Convolutional Neural Network DOI : 10.33633/joins.v7i2.7068 Nurul Huda, Sukmono Yogi Prayogi, Muna Adilah Ahmad, Alfa Yuliana Dewi	PDF 166-177
Sistem Informasi Manajemen Inventory Barang Gudang Berbasis Website Pada Bariklie Collection DOI : 10.33633/joins.v7i2.6888 Santi Oktafiani, Nur Hafifah Matondang, Rio Wirawan	PDF 178-189
Sistem Informasi Fee Streamer Berbasis Web Pada Shark Agency Dengan Pendekatan Extreme Programming DOI : 10.33633/joins.v7i2.6676 Lalang Erawan, Rizal Naufal Afif, Agus Winarno, Suharnawi Suharnawi	PDF 190-202

Editorial Team

Editor in Chief

[Sri Winarno, Ph.D.](#), Dian Nuswantoro University, Indonesia

Managing Editor

[Asih Rohmani](#), Universitas Dian Nuswantoro, Indonesia

Content Editor

[Ramadhan Rakhmat Sani](#), Universitas Dian Nuswantoro, Indonesia

Copy Editor

[Indra Gamayanto](#), Universitas Dian Nuswantoro, Indonesia

Layout Editor

[Ika Novita Dewi](#), Universitas Dian Nuswantoro, Indonesia

People

Reviewer

[Sendi Novianto](#), Dian Nuswantoro University, Indonesia

[Clara Hetty Primasari](#), Program Studi Sistem Informasi, Universitas Atma Jaya Yogyakarta, Indonesia

[Dr. Arta Moro Sundjaja](#), Bina Nusantara University, Indonesia

[Dr. ST Dwiwarso Utomo, SE, M.Kom, Akt, CA](#), Universitas Dian Nuswantoro, Indonesia

[Eko Hari Rachmawanto](#), Universitas Dian Nuswantoro, Indonesia

[Rizka Nugraha Pratikna](#), Universitas Katolik Parahyangan, Indonesia

[Aulia Akhrian Syahidi](#), Augmented Reality and Virtual Reality Laboratory, Politeknik Negeri Banjarmasin, Indonesia

[Mrs. Sucianna Ghadati Rabiha](#), Bina Nusantara University, Indonesia

[Dr. Yuventius Tyas Catur Pramudi, SSI, M.Kom](#), Universitas Dian Nuswantoro, Indonesia

[Dr Farrikh Alzami](#), Universitas Dian Nuswantoro Semarang, Indonesia

[Rangga Sidik](#), Universitas Komputer Indonesia, Indonesia

[Farah Zakiyah Rahmanti](#), Institut Teknologi Telkom Surabaya, Indonesia

[Septian Enggar Sukmana](#), Politeknik Negeri Malang, Indonesia

[Dr. Ruri Suko Basuki, M.Kom](#), Universitas Dian Nuswantoro, Indonesia

[Pujiono Pujiono](#)

[Mr. De Rosal Ignatius Moses Setiadi](#), Dian Nuswantoro University, Indonesia

[Christy Atika Sari](#), Dian Nuswantoro University, Indonesia

Klasifikasi Malaria Menggunakan Citra Sel Darah Merah Dengan Algoritma Convolutional Neural Network

Nurul Huda^{*1}, Sukmono Yogi Prayogi², Muna Adilah Ahmad³, Alfa Yuliana Dewi⁴

^{1,2,3}Fakultas Sains dan Teknologi, Institut Teknologi Statistika dan Bisnis (ITESA) Muhammadiyah Semarang

⁴Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan
e-mail:¹nurul.huda@itesa.ac.id, ²sukmono.yogi@itesa.ac.id, ³munaadilah123@gmail.com,

⁴alfayuliana@gmail.com

^{*}Penulis Korespondensi

Diterima: 27 Oktober 2022; Direvisi: 10 November 2022; Disetujui: 11 November 2022

Abstrak

Malaria merupakan infeksi penyakit yang ditularkan oleh gigitan nyamuk melalui protozoa parasit. Deteksi penyakit ini menggunakan pengambilan sampel sel darah merah dengan menghitung seberapa tinggi jumlah parasit di dalam sel darah merah. Penelitian untuk klasifikasi malaria telah banyak dilakukan, algoritma yang paling banyak digunakan yaitu Convolutional Neural Network (CNN). CNN merupakan algoritma Neural Network (NN) yang lazim digunakan untuk pengolahan citra. CNN mempunyai beberapa arsitektur, salah satunya adalah DenseNet, yang diusulkan dalam penelitian ini dikarenakan mampu menaikkan akurasi dari penelitian sebelumnya. Penelitian sebelumnya dengan dataset yang sama memilih arsitektur ResNet, GoogleNet, dan VGGNet untuk klasifikasi dan menghasilkan kisaran akurasi 90% - 96%, sedangkan penelitian terbaru menggunakan metode MM-Resnet menghasilkan akurasi 98%. Penelitian ini menggunakan dataset yang sama dengan penelitian sebelumnya yaitu dataset umum berupa citra sel darah merah yang terinfeksi dan tidak terinfeksi malaria. Eksperimen yang dilakukan menggunakan metode yang berbeda yaitu dengan melakukan pre-prosesing citra sel darah merah dengan cara resize gambar, augmentasi gambar, dan reduksi fitur gambar menggunakan metode Principal Component Analysis (PCA), hasil dari fitur reduksi tersebut selanjutnya diklasifikasikan menggunakan metode CNN. Arsitektur DenseNet menunjukkan akurasi terbaik yaitu 98,30%.

Kata kunci: image processing, CNN Classifier, Malaria, image augmentation, PCA

Abstract

Malaria is an infectious disease transmitted by mosquito bites through parasitic protozoa. Detection of this disease using red blood cell sampling by counting how high the number of parasites in the red blood cells is. Many studies have been conducted to classify malaria, and the most widely used algorithm is the Convolutional Neural Network (CNN). CNN is a Neural Network (NN) algorithm commonly used for image processing. CNN has several architectures, one of which is DenseNet, proposed in this study because it can increase the accuracy of previous research. Previous research chose ResNet, GoogleNet, and VGGNet architectures for classification, resulting in an accuracy range of 90% - 96%. However, the pre-processing image method only resizes the image. This study uses the same dataset as the previous research, namely the public dataset in images of red blood cells infected and not infected with malaria. Experiments were carried out using different methods, namely by pre-processing red blood cell images by resizing images, augmenting images, and reducing image features using the Principal Component Analysis (PCA) method. The dense-Net architecture shows the best accuracy of 98.30%.

Keywords: *image processing, CNN Classifier, Malaria, image augmentation, PCA*

1. PENDAHULUAN

Penelitian terkait malaria banyak dilakukan karena penyakit ini dapat berakibat fatal dan telah diinformasikan oleh *World Health Organization* (WHO) mencapai 219 juta kasus di seluruh dunia [1]. Malaria disebabkan oleh parasit *Plasmodium*, dengan sebagian besar yang terjadi pada manusia disebabkan oleh *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, dan *P. knowlesi*. *P. falciparum* dan *P. Vivax* [2]. Indonesia bagian timur merupakan wilayah endemi penyakit ini, sehingga penelitian terkait malaria lazim dilakukan guna menemukan metode yang tepat untuk mendeteksi penyakit ini guna mengurangi angka kematian yang disebabkan oleh malaria. Akhir-akhir ini dilakukan evaluasi terhadap kasus malaria di Timika untuk menemukan transmisi atau evolusi terhadap penyebaran malaria [3].

Machine Learning merupakan metode yang sering digunakan untuk otomatisasi deteksi suatu citra [4]. Kuantifikasi penularan malaria untuk klasifikasi risiko telah lama menjadi perhatian ahli epidemiologi [5]. Penelitian menggunakan algoritma *pre-trained deep convolutional neural network* menghasilkan arsitektur terbaik untuk mengklasifikasikan malaria yaitu *GoogleNet* dengan akurasi deteksi 93,89% dan sebagai pembanding yaitu arsitektur *ShuffleNet V2* memperoleh akurasi 95,20% dengan waktu tiga kali lebih cepat [6]. Klasifikasi gambar atau citra dapat dilakukan melalui banyak metode, salah satunya yaitu *image enhancement*, teknis sederhana yang mampu meningkatkan kontras gambar [7]. *Convolutional Neural Network* (CNN) juga pernah digunakan sebagai metode untuk mendeteksi malaria oleh Sammy V Militante dari Filipina, karena Filipina juga termasuk negara dengan kasus malaria tertinggi didunia selain Indonesia dan Vietnam. Arsitektur model CNN yang digunakan dalam penelitiannya yaitu *ResNet*, *GoogleNet*, dan *VGGNet*. Tahapan yang dilakukan dalam penelitian tersebut adalah *pre-processing image* dengan mengubah ukuran gambar menjadi 96x96 pixel, histogram, dan augmentatif. Pada tahapan pelatihan dilakukan ekstraksi fitur terhadap data dan menemukan arsitektur terbaik untuk klasifikasi malaria. Setelah model arsitektur ditentukan tahapan testing dan validasi dilakukan untuk menghitung akurasi deteksi [8]. Otomatisasi deteksi malaria berhasil mencapai akurasi terbaik dengan metode *Minimum Redudancy Maximum Relevance* (mRMR) dan PCA, karena metode ini mampu mengoptimalkan fitur, dengan cara mengurangi fitur dan parameter menjadi lebih sedikit [9].

Intelligent System saat ini juga banyak digunakan untuk membantu medis mengembangkan sistem lebih akurat dan lebih cepat, guna meminimalkan angka kematian pada pasien [10]. Rajaraman dan rekan melakukan penelitian untuk meningkatkan akurasi klasifikasi infeksi malaria dari data set sel darah merah yang terinfeksi protozoa dengan cara ekstraksi fitur dari gambar sel darah merah, kemudian mengklasifikasikan dengan metode CNN. Penelitian tersebut menghasilkan akurasi yang lebih tinggi dari penelitian sebelumnya yaitu 95,9% [11]. Selain itu *Multi-magnification deep residual neural network* (MM-ResNet) juga sempat diusulkan sebagai metode yang efektif yang dapat digunakan untuk otomatisasi klasifikasi gambar mikroskopis sel darah merah kedalam klasifikasi terinfeksi/tidak terinfeksi [12]. Pengelompokan data dengan *deep learning* arsitektur RBCNet juga pernah dilakukan untuk deteksi infeksi malaria dari sampel sel darah merah, menghasilkan akurasi 97% [13]. Sedangkan Kudisthalert mencoba menghitung dan mengklasifikasikan protozoa parasit penyebab infeksi malaria dengan metode AlexNet_FC7 dan menghasilkan akurasi 98.12% [14]. Dari penelitian dan sistem yang telah dikembangkan sebelumnya, masih menyisakan ruang untuk eksperimen ataupun perbaikan dengan cara mengombinasikan beberapa metode yang pernah dilakukan, seperti memilih metode *image processing* terbaik dan klasifikasi terbaik atau mencoba arsitektur lainnya dari CNN. Oleh karena itu, kami mengusulkan pendekatan baru dengan tujuan menghasilkan akurasi yang lebih baik.

2. METODE PENELITIAN

2.1 Pengumpulan Data

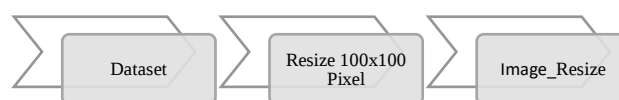
Dataset penelitian ini menggunakan dataset publik yang diambil dari Kaggle (<https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria>). Dataset terdiri dari 27.558 gambar yang diambil dari 150 pasien yang terinfeksi plasmodium falciparum di dalam sel darah merahnya dan 50 pasien sehat di Rumah Sakit Chittagong Medical College Bangladesh. Data kemudian diklasifikasikan secara manual oleh para ahli yang berada dalam Unit Penelitian Kedokteran Mahidol-Oxford, Bangkok, Thailand. Kumpulan data dikelompokkan menjadi dua yaitu terinfeksi dan tidak terinfeksi, masing-masing berjumlah 13.779 gambar. Alasan pemilihan dataset dikarenakan penelitian sebelumnya menggunakan dataset ini [6], [11], [15], [16] dan [17]. Ukuran gambar atau dimensi gambar untuk masing-masing sel darah merah terinfeksi dan tidak terinfeksi beragam. Dimensi gambar menggunakan satuan pixel, antara lain 139x139 pixel, 118x154 pixel, dan lain sebagainya, karena data beragam maka perlu dilakukan tahapan *pre-processing* terhadap data tersebut. Tahapan *pre-processing* gambar menggunakan kombinasi metode dari peneliti sebelumnya [9][18].

2.2 Tahap Pre-processing

Penelitian ini memerlukan pre-processing sebelum pemodelan, dikarenakan data set yang dimiliki belum terstruktur, ukuran gambar secara keseluruhan berbeda, hal ini akan mempengaruhi akurasi hasil pemodelan. Tahapan pre-processing ini bertujuan untuk mengubah data yang belum terstruktur menjadi data terstruktur sehingga analisis lanjutan untuk membangun sebuah model algoritma dapat dilakukan lebih baik. Berikut ini adalah tahapan pre-processing yang dilakukan penulis dalam penelitian ini:

1) Tahapan Resize Image

Dalam tahap ini, penulis mengubah ukuran data set gambar secara keseluruhan baik data set terinfeksi parasit maupun tidak terinfeksi menjadi 100x100 pixel dari total 227.558 data set. Gambar yang telah dirubah ukurannya disimpan ke dalam data set baru dengan nama "Image_Resize". Tahapan resize terdapat dalam Gambar 1.

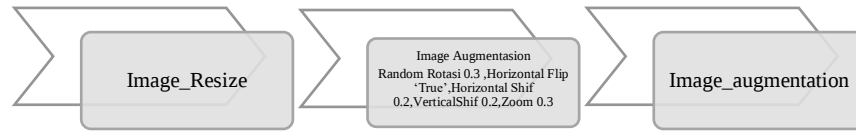


Gambar 1. Tahapan Resize Image

2) Tahapan Image Augmentation

Image augmentation yang digunakan adalah tradisional *Image augmentation* untuk memperluas data latih. Salah satu cara meningkatkan kinerja dan kemampuan model klasifikasi yaitu dengan memperbanyak data latih. *Image augmentation* dikenal memiliki performa yang baik khususnya untuk proses klasifikasi menggunakan metode CNN dengan jumlah data set kurang dari 1 juta. Di dalam penelitian ini data set yang digunakan sejumlah 13.779 gambar untuk masing-masing klasifikasi, untuk itu *Image augmentation* dipilih dengan tujuan memperluas data latih sehingga mampu menghasilkan akurasi yang lebih baik. *Image augmentation* mampu membentuk data latih baru secara otomatis dari data latih yang sudah ada. Tahapannya dapat dilihat pada Gambar 2. Transformasi gambar dalam tahapan augmentasi mencakup serangkaian operasi dari bidang manipulasi gambar, seperti shift, flips, zoom, dan banyak lagi. Dalam penelitian ini menggunakan

beberapa transformasi yaitu random rotasi, horizontal flip, vertical flip, horizontal shif, dan zoom.



Gambar 2. Tahapan Image Augmentation

3) Tahapan Fitur Reduction

Gambar digital biasa digambarkan menggunakan fungsi variabel $f(a,b)$, a dan b adalah koordinat spasial dan nilai $f(a,b)$ dari intensitas gambar pada koordinat itu sendiri. Terdapat dua jenis gambar yaitu *grayscale* dan RGB. Gambar *grayscale* terdiri dari dua komponen warna yaitu hitam dan putih, sedangkan RGB terdiri dari tiga warna dasar, yaitu merah, hijau, dan biru (*Red, Green, Blue*). Gambar atau citra digital RGB dengan ukuran 100x100 pixel artinya memiliki 30.000 fitur, hal tersebut dihitung dari 100x100x3 (3 adalah dimensi R, G, dan B). Jika *color encoding* dari gambar tersebut adalah 24 bit, maka dibutuhkan ruang penyimpanan sejumlah 100x100x24= 240.000 bit.

Tahapan fitur reduksi dibutuhkan untuk mengurangi jumlah fitur atau ruang dimensi dari sebuah gambar sehingga dapat mempercepat proses klasifikasi. Pengurangan dimensi dapat berdampak hilangnya informasi penting dari data asli, akan tetapi PCA mampu mempertahankan sebanyak mungkin informasi dari data asli dengan cara meminimalkan kesalahan. Untuk menentukan sub-ruang dimensi yang lebih kecil yang terbaik diperlukan *eigen value* dari matriks covarians x yang terbesar. Hal inilah yang disebut dengan istilah *Principal Component*. Tahapan seleksi fitur menggunakan metode PCA meliputi mencari nilai rata-rata pada persamaan (1), menghitung nilai zero mean pada persamaan (2), membangun matriks covarians pada persamaan (3) dan persamaan (4), menghitung *eigen value* pada persamaan (5), menghitung matriks *eigen vector* pada persamaan (6), dan mengurangi Jumlah dimensi N pada persamaan (7).

- a. Mencari nilai rata-rata (*mean*) dari data.

$$\underline{x} = \frac{x_1 + x_2 + \dots + x_m}{M} = \frac{\sum_{i=1}^M x_i}{M} \quad (1)$$

- b. Menghitung *zero mean* (masing-masing nilai dari data sampel dikurangi nilai rata-rata tiap parameter yang berhubungan).

$$\phi_i = x_i - \underline{x} \quad (2)$$

- c. Membangun matriks covarians dengan mengalikan matriks *zero mean* dengan transposenya

- Populasi

$$C = \frac{1}{N} \sum_{i=1}^M \phi_i^T \phi_i \quad (3)$$

- Sampel

$$C = \frac{1}{N-1} \sum_{i=1}^M \phi_i^T \phi_i \quad (4)$$

- d. Menghitung *eigen value* dari C .

$$\begin{aligned} C_u &= \lambda_u \quad ; \quad C_u - \lambda_u = 0 \quad ; \quad (C - \lambda I)_u = 0 \\ IC_u &= I \lambda_u \\ C_u &= \lambda I_u \\ (\lambda I - C)u &= 0 \\ \det \det (\lambda I - C) &= 0 \end{aligned}$$

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 & \lambda_2 & \dots & 0 & \dots & 0 & \dots & \dots & 0 & \dots & \lambda_n \\ -c_{1,1} & c_{1,2} & c_{2,1} & \dots & c_{m,1} & c_{2,2} & \dots & c_{m,2} & \dots & c_{1,n} & \dots & \dots & c_{2,n} & \dots & c_{m,n} \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 - c_{1,1} & -c_{1,2} & -c_{2,1} & \dots & -c_{m,1} & \lambda_2 - c_{2,2} & \dots & -c_{m,2} & \dots & -c_{1,n} & \dots & \dots & -c_{2,n} & \dots & \dots \\ & & & & & & & & & & & & & & \lambda_n - c_{m,n} \end{bmatrix}$$

$$\text{Hasil: } \lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n \quad (5)$$

e. Menghitung matriks *eigen vector*

- Dari *eigen value* yang dihitung pada langkah sebelumnya disubstitusikan ke dalam rumus:

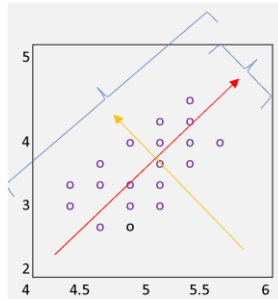
$$(\lambda I - C)u = 0 \quad (6)$$

- Selesaikan dengan menemukan nilai u
- Hasil: $u_1, u_2, u_3, \dots, u_n$

f. Mengurangi dimensi N sebesar K dimensi yang didapat dari *eigen value* yang paling besar sampai dengan yang paling kecil sebanyak K pertama.

- Pilih fitur sebanyak K berdasarkan nilai *eigen value* yang paling besar
 $\hat{X} - \underline{X} = \sum_{i=1}^K b_i u_i \text{ where } K \ll N \quad (7)$
- $\hat{X}$ merupakan hasil transformasi dari x

PCA memproyeksikan suatu arah data di mana data tersebut mempunyai varians yang tinggi. Arah proyeksi ditentukan oleh *eigen vector* dari matriks *covariance* yang nilai *eigen value*-nya paling besar. Besaran nilai dari *eigen value* tersebut adalah nilai varians data sepanjang arah dari *eigen vector* (garis lurus kuning dan merah). *Eigen value* tergambar dalam Gambar 3 dan langkah-langkahnya yang meliputi pemilihan nilai K pada Persamaan (8), penentuan vektor asal pada Persamaan (9), penerapan PCA untuk mengurangi kesalahan dari proses rekonstruksi data pada Persamaan (10), dan menyimpulkan nilai error pada Persamaan (11).



Gambar 3. *Eigen value* hasil perhitungan PCA

- Pemilihan nilai K menggunakan kriteria berikut:

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > \text{Threshold (contohnya 0.9 atau 0.95)} \quad (8)$$

- Contoh kasus di atas dapat diartikan dengan kita “menyiapkan” 90% atau 95% informasi dari data yang tersedia.
- Jika $K=N$, artinya “disiapkan” 100% dari data yang ada.
- Vektor asal x dapat dibangun lagi dengan prinsip komponen seperti rumus di bawah ini:

$$\hat{X} - \underline{X} = \sum_{i=1}^K b_i u_i \text{ atau } \hat{X} = \sum_{i=1}^K b_i u_i + \underline{X} \quad (9)$$

- PCA mengurangi kesalahan dari proses rekonstruksi data:

$$e = \|\underline{X} - \hat{X}\| \quad (10)$$

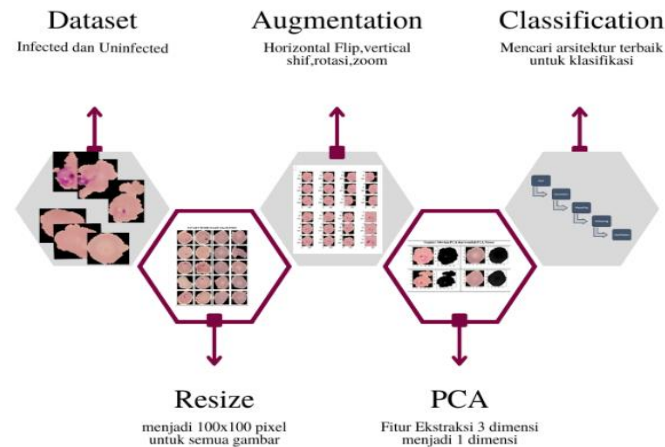
- Kesimpulannya eror sama dengan:

$$e = \frac{1}{2} \sum_{i=K+1}^N \lambda_i \quad (11)$$

2.3 Tahap Pemodelan

Gambar 4 menggambarkan keseluruhan model yang diusulkan dalam penelitian ini. Tujuan dari penelitian ini adalah mencari model algoritma terbaik untuk klasifikasi sel darah merah yang terinfeksi protozoa plasmodium dan sel darah merah tidak terinfeksi, sehingga ditemukan algoritma *machine learning* yang tepat untuk deteksi infeksi malaria. K-Fold Cross Validasi dikombinasikan dengan algoritma Convolutional Neural Network dikenal mampu mengoptimalkan klasifikasi [19].

Dalam penelitian ini, validasi menggunakan metode *K-Fold Cross Validation* dengan $K=5$, rancangan arsitektur CNN terdiri dari 4 layer konvolusi dan 1 layer *fully connected*. *K-Folds Validation* adalah suatu model atau metode statistika yang digunakan untuk melakukan evaluasi terhadap performa dari suatu algoritma. *5-Fold Cross Validation* sering direkomendasikan sebagai metode yang mampu memperkirakan kemampuan model algoritma. Dalam penelitian ini data dibagi menjadi 5 layer berukuran sama, data latih menggunakan 4-Fold dan 1-Fold terakhir digunakan untuk evaluasi data uji atau data validasi. Total data set 27.558 gambar 90% (24.800 gambar) data setnya akan digunakan untuk tahap *training*, dan 10% (2.758 gambar) digunakan untuk tahapan data testing. Dari 24.800 data set dibagi menjadi 5-Fold (masing-masing fold memiliki 4.960 gambar), 19.840 gambar sebagai data *training* dan 4.960 gambar sebagai data testing. Proses validasi *5-Fold Cross Validation* dapat dilihat pada Gambar 5.



Gambar 4 menunjukkan tahapan model yang diusulkan dalam penelitian ini, step 1 yaitu resize image, step 2 image augmentation, step 3 fitur reduksi dengan PCA, step terakhir classification

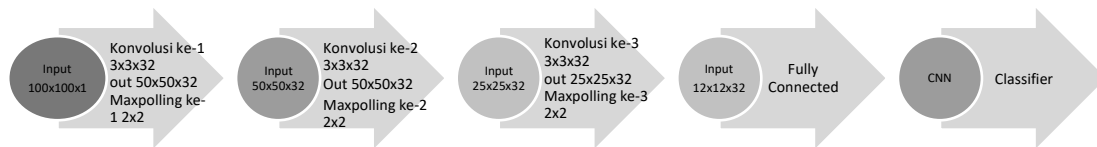
Gambar 4. Tahapan pemodelan yang diusulkan.

k-Fold	Cross Validation				
1	Validasi	training	training	training	training
2	training	Validasi	training	training	training
3	training	training	Validasi	training	training
4	training	training	training	Validasi	training
5	training	training	training	training	Validasi

Gambar 5. Tabel proses validasi menggunakan metode 5-fold cross validation.

Algoritma CNN dipilih dikarenakan algoritma ini memiliki akurasi yang lebih tinggi untuk identifikasi gambar sel darah merah yang terinfeksi malaria dan tidak terinfeksi [16]. Pada

beberapa penelitian, hasil eksperimen menunjukkan bahwa DenseNet mencapai hasil lebih baik daripada ResNet, hanya saja membutuhkan sumber daya komputasi sedikit lebih tinggi [20]. Oleh karena itu, peneliti memilih arsitektur DenseNet dalam penelitian ini. Sedangkan untuk mengurangi sumber daya komputasi, penulis melakukan reduksi fitur gambar dari 3 dimensi menjadi 1 dimensi, sehingga diperoleh model dengan akurasi lebih baik dan sumber daya komputasi lebih sedikit. Tahapan pemodelan CNN yang diusulkan dalam penelitian ini terlihat pada Gambar 6.



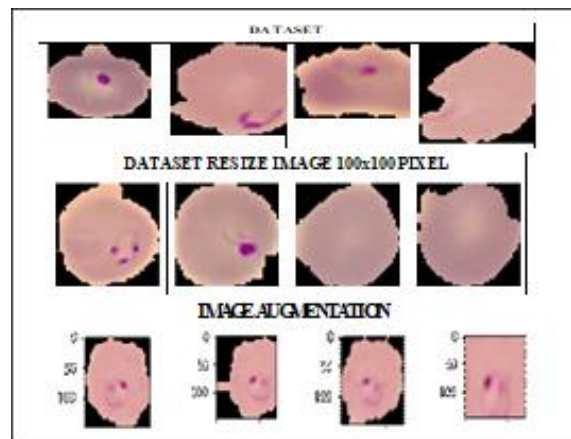
Gambar 6. Arsitektur CNN model yang diusulkan

3. HASIL DAN PEMBAHASAN

Bagian ini akan membahas tentang hasil penelitian yang telah dilakukan peneliti untuk mengetahui apakah hasil klasifikasi yang diusulkan dengan metode *image processing* dan klasifikasi *Convolutional Neural Network* (CNN) 5-Fold *Cross Validation* memiliki performa yang lebih baik dibandingkan dengan peneliti sebelumnya.

3.1. Tahapan *Pre-processing*

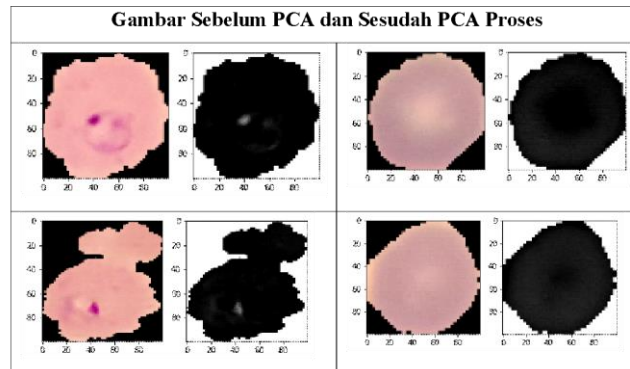
Pre-processing perlu dilakukan dikarenakan dataset memiliki ukuran gambar yang berbeda, hal ini akan mempengaruhi akurasi. Selain itu jumlah total data kurang dari 1 juta gambar, sehingga *image augmentation* dilakukan untuk memperbanyak dataset [16]. Tahapan pre-processing dapat dilihat pada Gambar 7.



Gambar 7. Tahapan preprocessing

3.2. Tahapan Fitur Reduksi dengan metode PCA

Tahapan terakhir dari *image preprocessing* yaitu reduksi fitur menggunakan metode PCA. Reduksi fitur dilakukan dengan tujuan untuk mempercepat proses komputasi. Dataset asli berupa gambar RGB dengan size 100x100 pixel, maka jumlah fitur untuk 1 gambar yaitu $100 \times 100 \times 3 = 30.000$ fitur. Untuk itu fitur reduksi perlu dilakukan dengan cara mengubah dimensi gambar menjadi 1 dimensi 100x100x1. Tampilan hasil eksperimen fitur reduksi dengan metode PCA tampak pada Gambar 8.



Gambar 8. Fitur Reduksi 3 dimensi menjadi 1 dimensi gambar

3.3. Tahapan Convolutional Neural Network Classifier

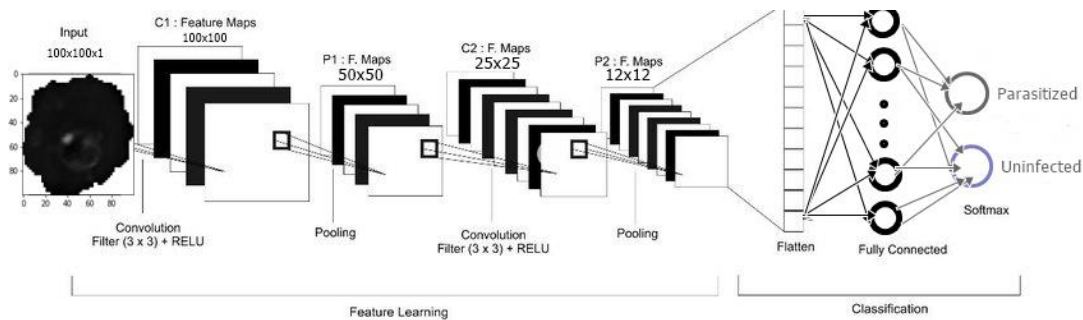
Arsitektur CNN Model dari penelitian ini yang digunakan adalah densenet-121, dikarenakan arsitektur ini memiliki performa yang paling tinggi. Tahapan layer dari arsitektur ini dalam dilihat pada Tabel 1.

Tabel 1. Tahapan Layer CNN Arsitektur Densenet-121

Tipe Layer	Bentuk Output	Parameter
conv2dim_1 (Conv2Dim)	(No, 100, 100, 32)	64
max_pooling2dim_1 (MaxPooling2)	(No, 50, 50, 32)	0
batch_normal_1 (Batch)	(No, 50, 50, 32)	128
DO_1 (Dropout)	(No, 50, 50, 32)	0
conv2dim_2 (Conv2Dim)	(No, 25, 25, 32)	1056
max_pooling2dim_2 (MaxPooling2)	(No, 25, 25, 32)	0
batch_normal_2 (Batch)	(No, 2, 25, 32)	128
DO_2 (Dropout)	(No, 25, 25, 32)	0
conv2dim_3 (Conv2Dim)	(No, 12, 12, 32)	1056
max_pooling2dim_3 (MaxPooling2)	No, 12, 12, 32)	0
batch_normal_3 (Batch)	(No, 12, 12, 32)	128
DO_3 (Dropout)	(No, 12, 12, 32)	0
flat_1 (Flatten)	(No, 4608)	0
densenet_1 (Dense)	(No, 512)	2359808
batch_normal_1 (Batch)	(No, 512)	2048
DO_1 (Dropout)	(No, 512)	0
densenet_2 (Dense)	(No, 2)	1026
<i>Total Parameter</i>		2,365,442
<i>Trainable Parameter</i>		2,364,226
<i>Non-Trainable Parameter</i>		1,216

Gambar 9 menunjukkan tahapan klasifikasi menggunakan metode CNN. Blok konstruksi utama CNN adalah konvolusi layer dan menerapkan filter yang dipilih ke gambar untuk menghapus fitur tingkat rendah dan tinggi pada gambar. Filter yang biasanya multidimensi dan masing-masing berisi nilai piksel yang mewakili tinggi-lebar matriks dan kedalaman matriks dapat menjadi filter yang mendeteksi tepi atau mendeteksi objek apa pun [17]. Pooling Layer yang sering ditambahkan di antara lapisan konvolusi berturut-turut adalah untuk mengurangi

ukuran pergeseran dan jumlah parameter dan perhitungan dalam jaringan. Tahapan terakhir yaitu flattening, pada tahap ini data masukan sudah siap untuk di klasifikasikan.



Gambar 9. Model klasifikasi gambar 1 dimensi menggunakan metode CNN

Dalam penelitian ini dilakukan eksperimen beberapa model klasifikasi untuk mendapatkan akurasi yang terbaik. 3 model klasifikasi yang diusulkan terdapat dalam Tabel 2, yaitu:

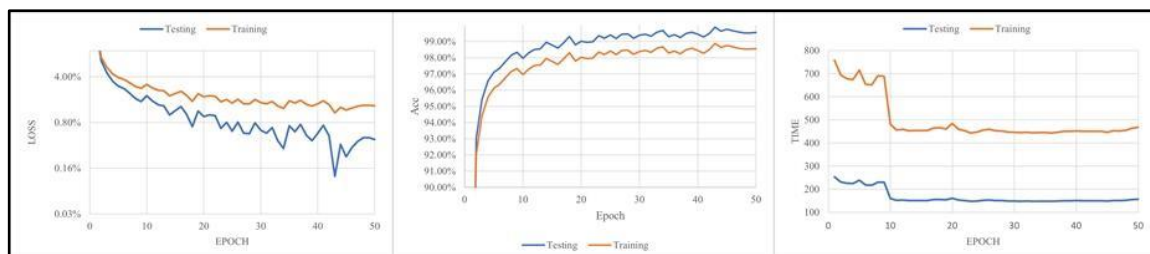
1. Metode Klasifikasi CNN tanpa *image augmentation* (No_Aug)
2. Metode Klasifikasi CNN dengan *image augmentation* (Aug)
3. Metode Klasifikasi CNN dengan *image augmentation* dan PCA fitur reduksi (Aug_PCA)

Tabel 2. Perbandingan Nilai Akurasi Ketiga Model

Model Klasifikasi	Akurasi		
	No_Aug	Aug	Aug_PCA
CNN Arsitektur DenseNet	95.51 %	97.30%	98.31%

3.4. Performa Convolutional Neural Network Classifier

Dari ketiga model klasifikasi diperoleh 1 model dengan akurasi terbaik yaitu CNN Classifier dengan *image augmentation* dan fitur reduksi PCA. Grafik perbandingan data training dan data testing yang meliputi loss, akurasi, dan time dapat dilihat pada Gambar 10. Grafik menunjukkan perbandingan loss training dan testing untuk epoch 1 sampai dengan 50. Akurasi data testing lebih tinggi dibandingkan dengan data training. Akurasi tertinggi terdapat pada epoch ke-43 yaitu 99,88% pada tahapan testing dan 98,88% pada tahapan training. Waktu komputasi tercepat terjadi pada epoch ke-23, 31, dan 33-35.

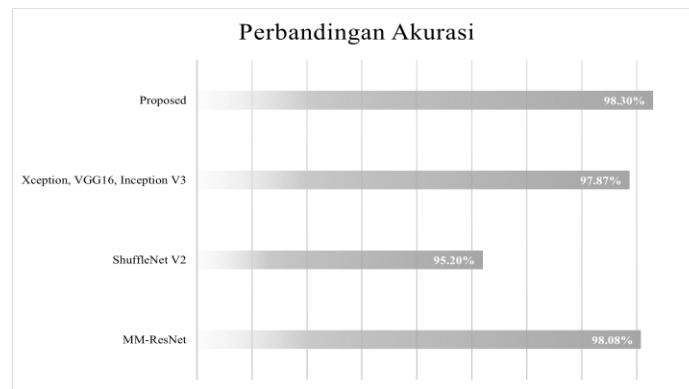


Gambar 10. Grafik Perbandingan *loss*, akurasi dan waktu komputasi. Gambar sebelah kiri menunjukkan grafik *loss*, gambar tengah grafik akurasi, dan sebelah kanan grafik waktu komputasi

3.5. Komparasi Klasifikasi Model

Dari hasil penelitian diperoleh model terbaik untuk mengklasifikasikan citra sel darah merah menjadi kelas terinfeksi dan tidak terinfeksi. Jika penelitian sebelumnya menggunakan

metode kombinasi ensambel arsitektur CNN, sehingga didapatkan kombinasi terbaik yaitu Xception, VGG19, dan InceptionV3 [16]. Selain itu arsitektur ShuffleNet V2 juga memiliki akurasi yang terbaik pada penelitian sebelumnya [6]. Penelitian terbaru yang dilakukan menggunakan metode MM-ResNet dan menghasilkan akurasi 98,08% [12]. Dari penelitian yang pernah dilakukan, peneliti menggabungkan metode terbaik untuk *image processing* dan juga arsitektur CNN, sehingga dihasilkan akurasi 98,30% dengan usulan metode *Image Augmentation*, PCA fitur reduksi, dan arsitektur DenseNet untuk klasifikasi. Grafik perbandingan tampak pada Gambar 11.



Gambar 11. Grafik Perbandingan Akurasi

4. KESIMPULAN

Tabel dan grafik dari eksperimen yang sudah ditampilkan dalam sub bab sebelumnya menggambarkan bahwa pemilihan fitur dan metode yang tepat mampu menaikkan nilai akurasi dari klasifikasi. Analisa terhadap dataset dan uji coba beberapa metode diperlukan untuk mencari model klasifikasi yang tepat untuk tipe dataset yang dimiliki. CNN sampai saat ini memiliki performa yang baik untuk klasifikasi, dikarenakan memiliki layer-layer konvolusi di dalam prosesnya dan dilakukan berulang-ulang untuk proses pembelajarannya [21]. Untuk tipe gambar R,G,B dengan nilai dimensi 3, membutuhkan proses reduksi dikarenakan untuk mempercepat proses komputasi. Reduksi fitur banyak metodenya, akan tetapi PCA mampu mereduksi fitur tanpa menghilangkan fitur aslinya, tetapi dengan cara menghitung dan mencari pemilihan fitur yang tepat dan mampu mentransformasi gambar dari dimensi 3 menjadi dimensi 1. Sedangkan *image augmentasi* bisa meningkatkan akurasi dikarenakan memperbanyak dataset, semakin banyak dataset maka proses pembelajaran semakin baik, sehingga *machine learning* yang dihasilkan optimal. Pemilihan arsitektur yang sama dengan tambahan proses *image processing* atau fitur reduksi mampu meningkatkan nilai akurasi untuk dataset di dalam penelitian ini.

Pada penelitian ini dapat disimpulkan penggunaan *Image Augmentation* dan *Principal Component Analysis* sangat signifikan dalam menaikkan akurasi, khususnya dalam pendeteksian objek dengan jumlah data set kurang dari 1juta data set untuk klasifikasi model menggunakan CNN, model dapat mengenali objek sel darah merah yang terinfeksi malaria (*protozoa plasmodium*) dan tidak terinfeksi, dan objek gambar hasil reduksi fitur menggunakan PCA mampu mempercepat proses klasifikasi dikarenakan memiliki dimensi lebih sedikit. Kesimpulan dari penelitian ini adalah model berhasil mendeteksi citra sel darah merah untuk kategori terinfeksi dan tidak terinfeksi. Penggunaan augmentasi untuk citra berhasil meningkatkan akurasi sebesar 1.79% karena data *training* menjadi lebih banyak. Fitur reduksi PCA berhasil menurunkan waktu komputasi dan meningkatkan akurasi sebesar 0.43%. penelitian ini menghasilkan klasifikasi model dengan akurasi 98.30% untuk mendeteksi penyakit malaria.

DAFTAR PUSTAKA

- [1] K. Manning, X. Zhai, and W. Yu, "Image analysis and machine learning-based malaria assessment system," *Digit. Commun. Networks*, vol. 8, no. 2, pp. 132–142, 2022.
- [2] R. J. Longley *et al.*, "Plasmodium vivax malaria serological exposure markers: Assessing the degree and implications of cross-reactivity with P. knowlesi," *Cell Reports Med.*, vol. 3, no. 6, p. 100662, 2022.
- [3] N. Lusiyanan dan A. Ahdika, "Evaluating recurrent episodes of malaria incidence in Timika, Indonesia, through a Markovian multiple-state model," *Infect. Dis. Model.*, vol. 7, no. 3, pp. 261–276, 2022.
- [4] A. Bin Abdul Qayyum, T. Islam, dan M. A. Haque, "Malaria Diagnosis with Dilated Convolutional Neural Network Based Image Analysis," *BECITHCON 2019 - 2019 IEEE Int. Conf. Biomed. Eng. Comput. Inf. Technol. Heal.*, pp. 68–72, 2019.
- [5] S. I. Hay, D. L. Smith, dan R. W. Snow, "Measuring malaria endemicity from intense to interrupted transmission," *Lancet Infect. Dis.*, vol. 8, no. 6, pp. 369–378, 2008.
- [6] I. G. S. M. Diyasa, A. Fauzi, A. Setiawan, M. Idhom, R. R. Wahid, dan A. D. Alhajir, "Pre-Trained Deep Convolutional Neural Network for Detecting Malaria on the Human Blood Smear Images," *3rd Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2021*, pp. 235–240, 2021.
- [7] J. Somasekar, G. Ramesh, G. Ramu, P. Dileep Kumar Reddy, B. Eswara Reddy, dan C. H. Lai, "A dataset for automatic contrast enhancement of microscopic malaria infected blood RGB images," *Data Br.*, vol. 27, p. 104643, 2019.
- [8] S. V. Militante, "Malaria Disease Recognition through Adaptive Deep Learning Models of Convolutional Neural Network," *ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci.*, 2019.
- [9] M. M. Hossain, M. A. Rahim, A. N. Bahar, dan M. M. Rahman, "Automatic malaria disease detection from blood cell images using the variational quantum circuit," *Informatics Med. Unlocked*, vol. 26, p. 100743, 2021.
- [10] K. E. D. Penas, P. T. Rivera, dan P. C. Naval, "Malaria Parasite Detection and Species Identification on Thin Blood Smears Using a Convolutional Neural Network," *Proc. - 2017 IEEE 2nd Int. Conf. Connect. Heal. Appl. Syst. Eng. Technol. CHASE 2017*, pp. 1–6, 2017.
- [11] S. Rajaraman *et al.*, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, 2018.
- [12] P. A. Pattanaik, M. Mittal, M. Z. Khan, dan S. N. Panda, "Malaria detection using deep residual networks with mobile microscopy," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1700–1705, 2022.
- [13] Y. M. Kassim *et al.*, "Clustering-Based Dual Deep Learning Architecture for Detecting Red Blood Cells in Malaria Diagnostic Smears," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 5, pp. 1735–1746, 2021.
- [14] W. Kudisthalert, K. Pasupa, dan S. Tongshima, "Counting and classification of malarial parasite from giemsa-stained thin film images," *IEEE Access*, vol. 8, pp. 78663–78682, 2020.
- [15] H. A. Nugroho, S. A. Akbar, dan E. E. H. Murhandarwati, "Feature extraction and classification for detection malaria parasites in thin blood smear," *ICITACEE 2015 - 2nd Int. Conf. Inf. Technol. Comput. Electr. Eng. Green Technol. Strength. Inf. Technol. Electr. Comput. Eng. Implementation, Proc.*, vol. 1, no. c, pp. 197–201, 2016.

- [16] H. K. Ragb, I. T. Dover, dan R. Ali, "Deep convolutional neural network ensemble for improved malaria parasite detection," *Proc. - Appl. Imag. Pattern Recognit. Work.*, vol. 2020–Octob, 2020.
- [17] S. Saglam, F. Tat, dan S. Bayar, "FPGA Implementation of CNN Algorithm for Detecting Malaria Diseased Blood Cells," *2019 Int. Symp. Adv. Electr. Commun. Technol. ISAECT 2019*, pp. 0–4, 2019.
- [18] M. Azeem, S. Javaid, H. Fahim, dan N. Saeed, "Neural Networks for Infectious Diseases Detection: Prospects and Challenges," *arXiv Prepr. arXiv2112.03571*, pp. 0–16, 2021.
- [19] Y. N. Fuadah, I. D. Ubaidullah, N. Ibrahim, F. F. Taliningsing, N. K. Sy, dan M. A. Pramuditho, "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 3, p. 728, 2022.
- [20] C. Zhang *et al.*, "ResNet or DenseNet? Introducing dense shortcuts to ResNet," *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, pp. 3549–3558, 2021.
- [21] B. S. Gandhi, S. A. U. Rahman, A. Butar, dan A. Victor, "Brain tumor segmentation and detection in magnetic resonance imaging (MRI) using convolutional neural network," *Brain Tumor MRI Image Segmentation Using Deep Learn. Tech.*, pp. 37–57, Jan. 2021.