Analisis Fitur Warna untuk Klasifikasi Tingkat Keparahan Kanker Serviks

1st Maya Yolanda
Jurusan Sistem Komputer
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia

2nd Muhammad Naufal Rachmatullah*
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
naufalrachmatullah@gmail.com

3rd Akhiar Wisata Arum
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
wistaarum16@gmail.com

Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
yolandamaya929@gmail.com

4th Ade Iriani Sapitri
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
adeirianisapitri13@gmail.com

5th Anggun Islami
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
anggunislami2@gmail.com

*Corresponding author: naufalrachmatullah@gmail.com

***Abstrak*—Kanker serviks merupakan salah satu penyebab utama kematian perempuan secara global, sehingga deteksi dini sangat penting untuk menurunkan angka mortalitas. Penelitian ini mengevaluasi peran *feature engineering* berbasis warna pada citra kolposkopi dalam klasifikasi tingkat keparahan lesi serviks. Tiga ruang warna digunakan untuk ekstraksi fitur, yaitu *Red-Green-Blue*, *Hue-Saturation-Value*, dan CIELAB, pada citra tanpa segmentasi *Region of Interest*. Fitur yang diperoleh digunakan sebagai masukan pada tiga algoritma klasifikasi: *Multi-Layer Perceptron*, *Support Vector Machine*, dan *Extreme Gradient Boosting*. Evaluasi dilakukan menggunakan *confusion matrix* untuk menghitung akurasi, presisi, sensitivitas, dan F1-Score. Hasil menunjukkan bahwa penggabungan ketiga ruang warna memberikan performa terbaik, dengan *Support Vector Machine* mencapai akurasi 66% dan F1-Score 54%. Temuan ini menegaskan pentingnya representasi warna dalam meningkatkan kinerja klasifikasi lesi serviks dan mendukung pengembangan sistem pendukung keputusan berbasis komputer yang lebih akurat dan efisien.**

***Keywords*—Kanker serviks, Fitur warna, *Machine Learning*, Klasifikasi.**

I. PENDAHULUAN

Kanker serviks merupakan penyebab utama kematian pada perempuan di dunia, dengan sekitar 604.000 kasus baru dan 342.000 kematian per tahun menurut *GLOBOCAN 2020*[1]. Di Indonesia, penyakit ini menempati posisi kedua setelah kanker payudara, dengan lebih dari 36.000 kasus baru dan 21.000 kematian setiap tahunnya [2]. Sebagian besar kasus ini disebabkan oleh infeksi *Human Papillomavirus* (HPV) [3]. Tingginya angka kematian mencerminkan keterlambatan diagnosis akibat keterbatasan fasilitas dan tenaga ahli[4], [5].

Kemajuan *Artificial Intelligence* (AI) dan *Computer-Aided Diagnosis* (CAD) telah meningkatkan akurasi deteksi kanker serviks [6]. Berbagai penelitian sebelumnya telah menggunakan metode berbasis *deep learning* maupun *machine learning* untuk klasifikasi citra serviks, seperti Ma et al. [7] mengembangkan sistem CAD berbasis *deep convolutional network* untuk mengidentifikasi tingkat displasia serviks dari citra kolposkopi. Pendekatan tersebut belum mengoptimalkan informasi warna yang penting dalam representasi morfologi jaringan. Selanjutnya, Rahaman et al. [8] mengusulkan *DeepCervix*, framework berbasis *hybrid deep feature fusion*, yang berhasil meningkatkan akurasi klasifikasi, tetapi prosesnya bergantung pada ekstraksi fitur kompleks yang memerlukan sumber daya komputasi tinggi. Sementara itu, metode berbasis SVM juga menunjukkan hasil baik, tetapi belum memanfaatkan potensi ruang warna secara mendalam [4], [8].

Keterbatasan penelitian-penelitian terdahulu menunjukkan bahwa pemanfaatan informasi warna pada citra kolposkopi masih belum dioptimalkan secara sistematis[9] Warna merupakan salah satu indikator penting dalam menggambarkan tingkat keparahan lesi serviks, sehingga analisis berbasis warna berpotensi meningkatkan keakuratan klasifikasi [10]. Berdasarkan hal tersebut, penelitian ini berfokus pada *feature engineering* berbasis warna dengan mengeksplorasi tiga ruang warna utama, yaitu *Red-Green-Blue* (RGB), *Hue-Saturation-Value* (HSV), dan CIELab, untuk mendukung proses klasifikasi tingkat keparahan lesi serviks.

Penelitian ini bertujuan untuk membangun model klasifikasi citra kolposkopi berbasis *machine learning* menggunakan tiga algoritma, yaitu *Multi-Layer Perceptron* (MLP), *Support Vector Machine* (SVM), dan *Extreme*

Gradient Boosting (XGBoost). Pendekatan yang diusulkan memberikan kontribusi baru dengan menganalisis perbandingan efektivitas representasi warna terhadap performa model klasifikasi, serta memperkuat potensi pengembangan sistem *Computer-Aided Diagnosis (CAD)* yang lebih efisien, interpretable, dan adaptif terhadap variasi warna citra medis.

II. METODOLOGI

A. Ekstraksi Fitur Warna

Karakteristik visual citra serviks dianalisis melalui distribusi intensitas warna pada setiap kanal citra [11], karena variasi warna jaringan serviks merefleksikan perbedaan morfologi yang relevan untuk klasifikasi tingkat keparahan lesi [12]. Seluruh citra diubah ukurannya menjadi 224×224 piksel untuk menyeragamkan resolusi, kemudian dihitung histogram intensitas pada setiap kanal warna dengan 256 bin yang dinormalisasi sesuai Persamaan (1).

$$h_{norm}(i) = \frac{h(i)}{\sum_{j=0}^{N-1} h(j)}, \quad i = 0,1,\ldots,N-1 \tag{1}$$

Persamaan tersebut menggambarkan proses normalisasi histogram, dimana *h(i)* menyatakan jumlah piksel pada bin ke-*i*, dan N adalah jumlah total bin. Normalisasi ini membuat nilai fitur merepresentasikan distribusi relatif intensitas warna, bukan jumlah absolut, sehingga lebih robust terhadap variasi ukuran citra.

Ekstraksi fitur dilakukan pada tiga ruang warna, yaitu RGB, HSV, dan CIELab. Masing-masing ruang warna memiliki karakteristik berbeda. Ruang warna RGB merepresentasikan warna berdasarkan intensitas merah (R), hijau (G), dan biru (B) sebagaimana dinyatakan pada Persamaan (2). Ruang warna HSV didasarkan pada atribut hue (H), saturation (S), dan value (V) seperti pada Persamaan (3). Sementara itu, ruang warna CIELab menggambarkan persepsi visual manusia terhadap kecerahan dan kontras warna sebagaimana dinyatakan pada Persamaan (4).

$$F_{RGB=}[h^R_{norm,}, h^G_{norm,}, h^B_{norm,}] \tag{2}$$
$$F_{HSV=}[h^H_{norm,}, h^S_{norm,}, h^V_{norm,}] \tag{3}$$
$$F_{Lab=}[h^L_{norm,}, h^a_{norm,}, h^b_{norm,}] \tag{4}$$

Persamaan (2)–(4) merepresentasikan vektor fitur dari ruang warna RGB, HSV, dan CIELab yang masing-masing dibentuk dari nilai normalisasi intensitas komponennya: merah–hijau–biru (R, G, B), *hue–saturation–value* (H, S, V), serta kecerahan dan sumbu warna (L*, a*, b*).

Untuk representasi gabungan, vektor fitur dari ketiga ruang warna dikonkatenasi menjadi satu vektor dinyatakan pada persamaan 5.

$$[F_{All=} F_{RGB,} F_{HSV,} F_{Lab}] \tag{5}$$

Selain itu, ekstraksi dilakukan melalui dua pendekatan, yaitu menggunakan citra utuh (*non-segmented*) dan citra tersegmentasi berbasis *Region of Interest (ROI)*. Kedua hasil ekstraksi ini digunakan sebagai masukan pada model klasifikasi, sehingga memungkinkan evaluasi perbandingan pengaruh segmentasi terhadap kinerja sistem.

B. Metode Klasifikasi

Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) merupakan jenis jaringan saraf tiruan berlapis yang banyak digunakan untuk klasifikasi[13]. Arsitekturnya terdiri atas lapisan input, beberapa *hidden layer*, dan lapisan output. Setiap neuron pada lapisan ke-*l* menghitung aktivasi dari kombinasi linear bobot dan bias menggunakan fungsi aktivasi non-linear, seperti ReLU atau Sigmoid,sebagaimana ditunjukkan pada persamaan 6.

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$$
$$a^{(l)} = f(z^{(l)}) \tag{6}$$

Fungsi *loss* yang digunakan untuk pelatihan adalah *Categorical Cross-Entropy*, yang dirumuskan sebagai persamaan 7.

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{7}$$

Parameter jaringan diperbarui menggunakan algoritma backpropagation dengan optimisasi gradient descent untuk meminimalkan nilai loss. Struktur MLP yang digunakan pada penelitian ini ditunjukkan pada Gambar 1, mencakup lapisan input yang menerima vektor fitur warna hasil ekstraksi, beberapa hidden layer dengan sejumlah neuron, serta lapisan output yang merepresentasikan prediksi kelas citra serviks. Setiap neuron pada lapisan tersembunyi menggunakan fungsi aktivasi ReLU untuk memperkenalkan non-linearitas, sehingga jaringan mampu mempelajari pola kompleks pada data citra secara lebih efektif.

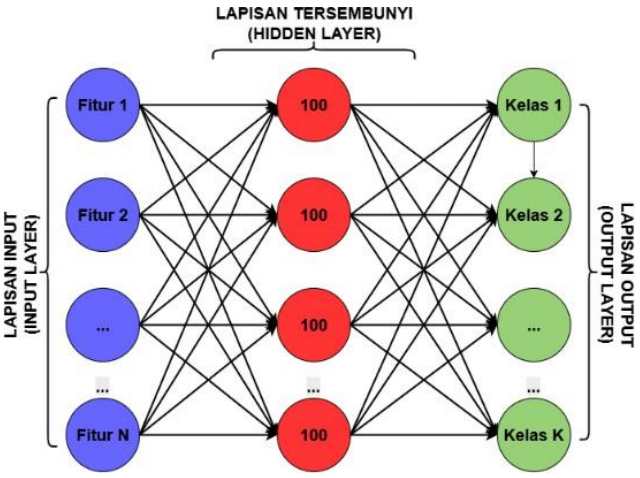


Fig. 1. Arsitektur *Multi-Layer Perceptron* (MLP)

Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang banyak digunakan karena kemampuannya dalam menangani data berdimensi tinggi. Prinsip utama SVM adalah mencari *hyperplane* pemisah (8)

yang mampu memaksimalkan margin antara dua kelas pada ruang fitur[14]. Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi *kernel* untuk memetakan data ke ruang berdimensi lebih tinggi agar lebih mudah dipisahkan. Fungsi keputusan SVM secara umum dinyatakan pada Persamaan 8.

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right)$$

Fungsi *kernel* yang digunakan dalam penelitian ini adalah *Radial Basis Function* (RBF), sebagaimana Persamaan 9.

$$K(x_i, x_j) = \exp\left(-\gamma ||x_i - x_j||^2\right) \quad (9)$$

(11)

Parameter γ mengatur luas pengaruh setiap data latih terhadap titik lain. Dengan pendekatan ini, SVM dapat membentuk *non-linear decision boundary* yang sesuai dengan distribusi fitur warna pada dataset.

Untuk klasifikasi multi-kelas, digunakan strategi one-vs-one, di mana SVM membangun beberapa model biner dan hasil akhir ditentukan berdasarkan mayoritas keputusan. Struktur SVM yang digunakan ditunjukkan pada Gambar 2, yang menggambarkan hyperplane, margin maksimum, dan support vectors sebagai penentu batas klasifikasi. Kernel RBF digunakan untuk memetakan data ke ruang berdimensi lebih tinggi sehingga model dapat memisahkan kelas yang tidak terpisahkan secara linear.

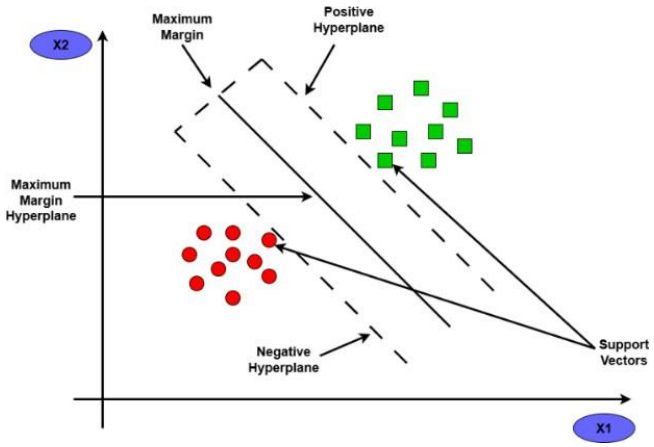


Fig. 2. Ilustrasi *Support Vektor Machine* (SVM)

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) merupakan algoritma klasifikasi berbasis *ensemble learning* yang menggabungkan sejumlah pohon keputusan (*decision trees*) secara bertahap untuk membentuk model prediksi yang lebih kuat[15]. XGBoost termasuk dalam kategori *gradient boosting*, di mana setiap pohon baru yang dibangun berfungsi untuk memperbaiki kesalahan prediksi dari pohon sebelumnya dengan cara meminimalkan fungsi loss melalui optimisasi berbasis gradien.

Secara umum, model prediksi XGBoost untuk data input x_i dapat dituliskan sebagai persamaan 10.

$$\hat{y}_\iota = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \quad (10)$$

Persamaan ini menggunakan $\mathcal{F}$ sebagai himpunan semua pohon keputusan, dan K adalah jumlah pohon yang dibangun.

Fungsi objektif fungsi yang diminimalkan pada XGBoost terdiri atas dua komponen, yaitu fungsi loss dan regularisasi, yang dituliskan sebagai persamaan 11.

$$L(\emptyset) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Persamaan ini menggunakan $l(y_i, \hat{y}_i)$ sebagai fungsi loss (misalnya log-loss untuk klasifikasi), sedangkan $\Omega(f_k)$ adalah fungsi regularisasi yang mengontrol kompleksitas pohon, sehingga mencegah overfitting.

Arsitektur XGBoost dapat divisualisasikan pada Gambar 3, yang memperlihatkan sekumpulan pohon keputusan yang dibangun secara berurutan. Setiap pohon berkontribusi dalam memperbaiki prediksi berdasarkan error dari pohon sebelumnya, hingga akhirnya menghasilkan prediksi akhir yang lebih akurat. Model ini juga menerapkan teknik regularisasi untuk menghindari overfitting dan meningkatkan kemampuan generalisasi terhadap data baru.

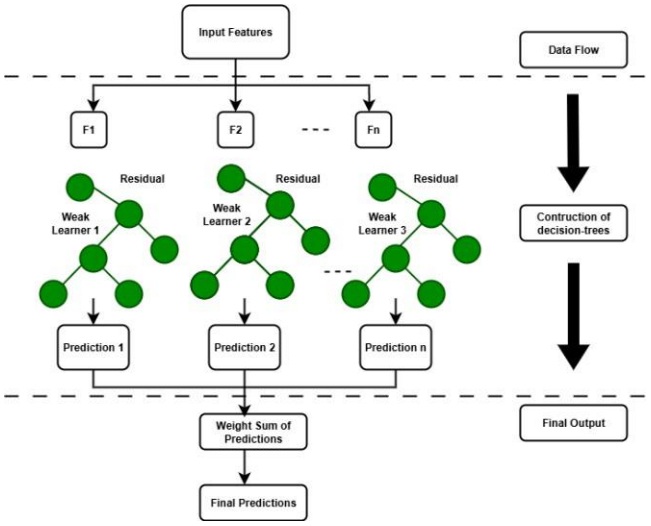


Fig. 3. Arsitektur XGBoost

III. EKSPERIMEN DAN HASIL

Bagian ini menjelaskan alur eksperimen yang meliputi pengambilan dan pra-pemrosesan citra sel serviks, ekstraksi fitur warna untuk membentuk representasi numerik, serta penerapan model klasifikasi MLP, SVM, dan XGBoost. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-Score yang dihitung berdasarkan

confusion matrix untuk menilai kemampuan model dalam mengklasifikasikan citra secara akurat.

A. Material dan Metode

Penelitian ini menganalisis dan mengklasifikasikan citra serviks untuk mendeteksi tingkat keparahan lesi berdasarkan ekstraksi fitur warna. Prosesnya mencakup pengumpulan, pra-pemrosesan, segmentasi, dan ekstraksi fitur warna guna memastikan kualitas serta konsistensi data sebelum klasifikasi. Seperti pada Gambar 4, penelitian dibagi menjadi dua pipeline: Pipeline 1 mengekstraksi fitur langsung dari citra asli, sedangkan Pipeline 2 menerapkan segmentasi *Region of Interest* (ROI) sebelum ekstraksi. Kedua pipeline ini dibandingkan untuk menilai pengaruh segmentasi terhadap kualitas fitur dan kinerja model klasifikasi secara menyeluruh dan terukur. Hasil perbandingan dari kedua pipeline tersebut diharapkan dapat memberikan gambaran yang lebih jelas mengenai efektivitas segmentasi dalam meningkatkan akurasi deteksi dan klasifikasi tingkat keparahan lesi serviks. Dengan demikian, rancangan ini membantu mengevaluasi kontribusi tiap tahap pemrosesan terhadap hasil akhir klasifikasi.

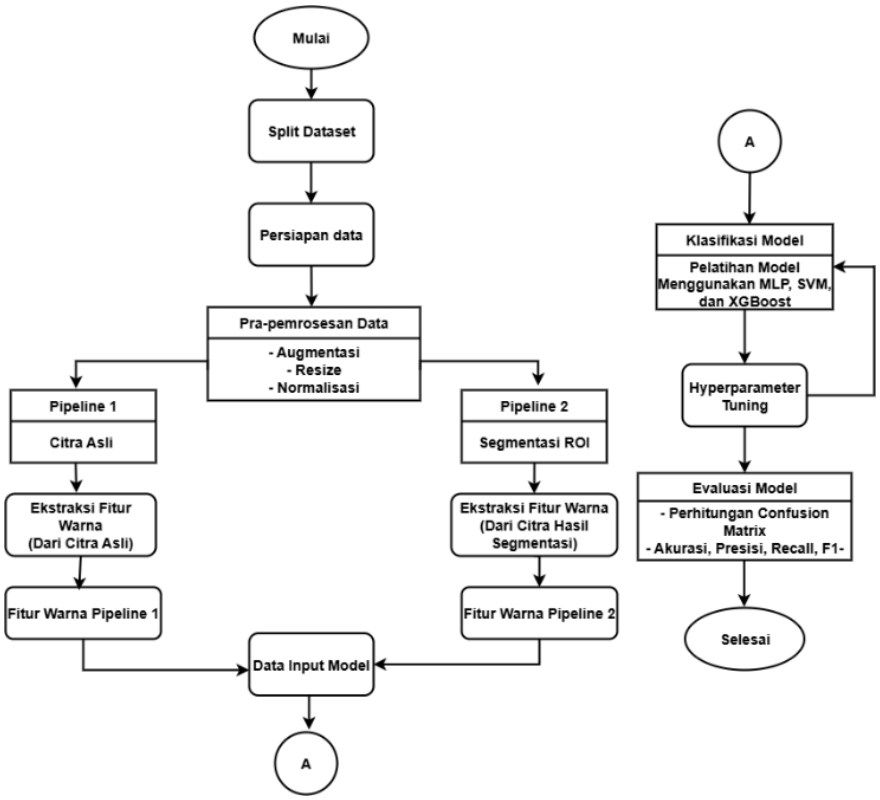


Fig. 4. Alur Tahapan Penelitian

Persiapan Data

Dataset penelitian ini berasal dari IARC Image Bank Colposcopy yang berisi tiga kategori visual epitel serviks (Transparent, Milky, Opaque) dalam kondisi *after acetic acid* [16]. Distribusi jumlah kasus dan citra untuk masing-masing kategori disajikan pada Tabel I, yang menunjukkan bahwa kategori *Opaque* memiliki jumlah sampel terbanyak dibandingkan dua kategori lainnya. Contoh citra representatif dari ketiga kategori tersebut ditampilkan pada Gambar 5, yang memperlihatkan perbedaan visual pada jaringan epitel serviks setiap kelas.

TABLE I. Distribusi Jumlah Case dan Citra Sebelum Augmentasi

Kategori	Jumlah Case	Jumlah Gambar
Transparent	86	118
Milky	35	48
Opaque	76	121
Total	**197**	**279**



(a) (b) (c)

Fig. 5. Citra Kanker Serviks

Pembagian dataset dilakukan berbasis *case* (pasien) untuk menghindari *data leakage* antara subset latih dan validasi. Proses *stratified split* dengan rasio 80:20 digunakan untuk menjaga proporsi kelas tetap seimbang pada kedua subset. Metode ini membagi data berdasarkan distribusi kelas yang

telah dihitung sebelumnya sehingga representasi tiap kategori tetap konsisten [17], [18]. Pendekatan ini mengikuti rekomendasi Johnson dan Khoshgoftaar [17] untuk penanganan ketidakseimbangan kelas serta implementasi oleh Pedregosa *et al.* [18] pada *scikit-learn* yang memastikan kestabilan performa model.

Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menyeragamkan ukuran, kualitas, dan distribusi citra sebelum ekstraksi fitur warna. Proses dimulai dengan augmentasi pada data pelatihan melalui rotasi, pergeseran, *zoom*, dan *horizontal flip* guna menyeimbangkan jumlah citra tiap kelas, sementara data uji dibiarkan tanpa augmentasi agar evaluasi tetap objektif. Untuk memastikan keseimbangan distribusi kelas, proses augmentasi dilakukan hingga total data latih mencapai 388 citra, terdiri atas 129 *Transparent*, 113 *Milky*, dan 105 *Opaque*. Selanjutnya, seluruh citra diubah ukurannya menjadi 224×224 piksel dan dinormalisasi menggunakan *Z-score normalization* dengan mengurangi rata-rata dan membagi dengan standar deviasi tiap citra. Tahapan ini memastikan keseimbangan intensitas antar-kanal warna dan menghasilkan data yang siap digunakan untuk ekstraksi fitur sebagaimana ditunjukkan pada Gambar 6.

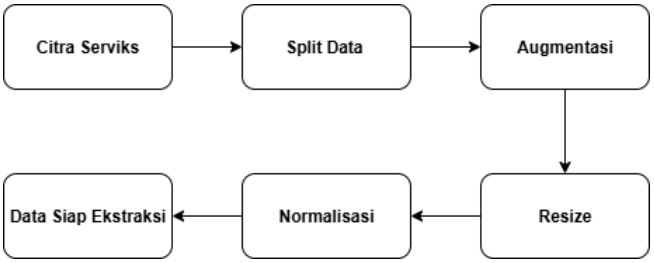


Fig. 6. Diagram Alur Pra-Pemrosesan Data

B. Ekstraksi Fitur

Proses ekstraksi fitur warna dilakukan melalui dua pipeline: Pipeline 1 mengekstraksi fitur dari citra asli, sedangkan Pipeline 2 mengekstraksi fitur dari citra setelah segmentasi ROI. Setiap citra diubah menjadi 224×224 piksel untuk menjaga konsistensi. Pada Pipeline 2, segmentasi ROI menyoroti area lesi serviks agar ekstraksi fokus pada region signifikan. Tiap kanal warna diubah menjadi histogram ter-normalisasi (min–max) untuk menyamakan skala intensitas. Gambar 7 menunjukkan alur ekstraksi, mulai dari pembacaan citra, normalisasi, ekstraksi histogram pada ruang warna RGB, HSV, dan CIELAB, hingga penggabungan vektor fitur. Hasil akhir berupa vektor fitur gabungan yang merepresentasikan karakteristik warna citra dan digunakan sebagai input klasifikasi tingkat keparahan lesi serviks.

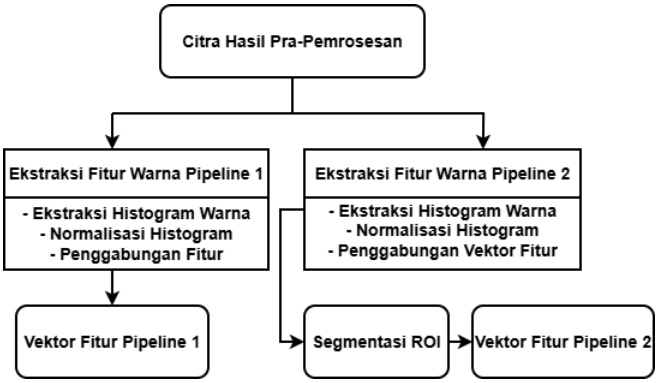


Fig. 7. Diagram Alur Ekstraksi Fitur Warna

Untuk memvisualisasikan distribusi fitur warna, Gambar 8 menampilkan histogram intensitas pada ruang warna RGB, HSV, dan CIELAB. Pada RGB, kanal R, G, dan B merepresentasikan distribusi linear warna primer; HSV menyoroti hue, saturasi, dan value yang menggambarkan rona, kontras, dan kecerahan; sedangkan CIELAB mencerminkan persepsi warna manusia melalui komponen L (kecerahan), a (merah–hijau), dan b (kuning–biru) dengan sebaran lebih merata. Perbedaan karakteristik ini memperkaya representasi citra dan meningkatkan akurasi klasifikasi lesi serviks.

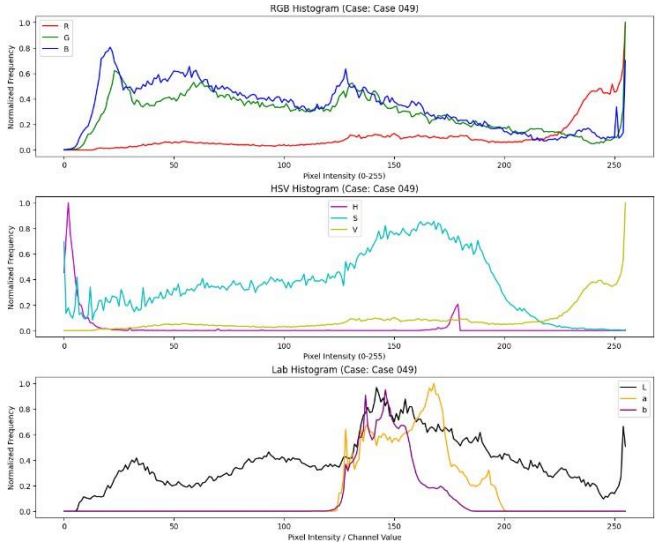


Fig. 8. Histogram Warna RGB, HSV,CIELAB

C. Hasil Evaluasi Model

Evaluasi dilakukan untuk menilai kinerja MLP, SVM, dan XGBoost dalam mengklasifikasikan tingkat keparahan lesi serviks berdasarkan fitur warna citra colposcopy. Implementasi menggunakan Python dengan TensorFlow/Keras untuk MLP serta *scikit-learn* untuk SVM dan XGBoost. Kinerja diukur menggunakan *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari *confusion matrix*. Tabel II menyajikan hasil perbandingan dua pipeline, yaitu Pipeline 1 (tanpa segmentasi ROI) dan Pipeline 2 (dengan segmentasi ROI). SVM dengan fitur gabungan RGB–HSV–CIELAB pada Pipeline 1 mencapai hasil terbaik (Acc = 0,66; F1 = 0,54), MLP meningkat pada Pipeline 2 dengan fitur RGB

(Acc = 0,61), sedangkan XGBoost relatif stabil pada kedua pipeline (HSV, F1 = 0,51). Parameter model dioptimasi melalui penyesuaian *C*, *gamma*, dan *kernel RBF* pada SVM; *hidden layer*, *activation*, *solver*, *learning rate* (0,001–0,01), *batch size* (32–64), dan *max_iter* (500–1000) pada MLP; serta *n_estimators* = 100, *max_depth* = 5, dan *learning_rate* = 0,1 pada XGBoost. Hasil menunjukkan bahwa kinerja model dipengaruhi oleh fitur warna, strategi penskalaan, dan parameter tuning, dengan Pipeline 1 unggul pada konteks global dan Pipeline 2 pada detail lokal.

TABLE II. Perbandingan Hasil Pipeline 1 dan Pipeline 2

Model	Warna	Acc (P1,%)	Prec (P1,%)	Rec (P1,%)	F1 (P1,%)	Acc (P2,%)	Prec (P2,%)	Rec (P2,%)	F1 (P2,%)
MLP	RGB	0.49	0.31	0.38	0.31	0.61	0.58	0.54	0.55
	HSV	0.56	0.38	0.45	0.41	0.51	0.34	0.41	0.37
	CIELAB	0.51	0.51	0.44	0.44	0.54	0.46	0.46	0.45
	Gabungan	0.61	0.40	0.48	0.42	0.56	0.47	0.56	0.51
SVM	RGB	0.44	0.28	0.34	0.30	0.56	0.54	0.56	0.54
	HSV	0.63	0.46	0.50	0.45	0.57	0.54	0.54	0.52
	CIELAB	0.61	0.41	0.48	0.43	0.56	0.63	0.56	0.52
	Gabungan	**0.66**	**0.61**	**0.55**	**0.54**	**0.54**	**0.53**	**0.54**	**0.51**
XGBOOST	RGB	0.56	0.49	0.48	0.47	0.51	0.44	0.51	0.47
	HSV	0.61	0.50	0.52	0.51	0.56	0.48	0.56	0.51
	CIELAB	0.56	0.49	0.45	0.42	0.49	0.43	0.49	0.45
	Gabungan	0.59	0.51	0.47	0.54	0.59	0.49	0.59	0.53

IV. DISKUSI

Analisis menunjukkan bahwa kombinasi ruang warna RGB, HSV, dan CIELAB memberikan kinerja terbaik dalam klasifikasi tingkat keparahan lesi serviks, dengan SVM pada Pipeline 1 (citra asli) mencapai akurasi 0,66 dan F1-Score 0,54. Ruang warna RGB menjaga kestabilan intensitas, HSV menangkap variasi hue akibat reaksi asam asetat, sedangkan CIELAB memperkuat kontras visual penting. Pipeline 2 (ROI disegmentasi) meningkatkan performa MLP (RGB, 0,61) namun relatif stabil pada SVM dan XGBoost, menunjukkan bahwa segmentasi ROI lebih efektif untuk model yang sensitif terhadap informasi lokal. Performa dipengaruhi oleh karakteristik algoritma, konteks spasial, penskalaan, dan tuning parameter. Studi ini menegaskan efektivitas feature engineering berbasis ruang warna serta peran segmentasi ROI sebagai pelengkap. Keterbatasan mencakup ukuran data dan kualitas citra; penelitian lanjutan disarankan mengintegrasikan segmentasi otomatis, fitur tekstur atau morfologi, dan dataset lebih besar untuk meningkatkan generalisasi model.

V. KESIMPULAN

Feature engineering berbasis ruang warna RGB, HSV, dan CIELAB terbukti efektif dalam klasifikasi tingkat keparahan lesi serviks, dengan kombinasi ketiganya menghasilkan performa paling stabil. Segmentasi ROI dapat meningkatkan sensitivitas model tertentu, namun efektivitasnya bergantung pada algoritma dan representasi warna. Hasil ini menegaskan kontribusi analisis ruang warna untuk meningkatkan akurasi model dan membuka peluang pengembangan metode lanjutan, seperti optimasi segmentasi ROI atau integrasi fitur tekstur.

REFERENSI

[1] H. Sung *et al*., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] B. A. Tjokroprawiro *et al*., "The challenging journey of cervical cancer diagnosis and treatment at the second largest hospital in Indonesia," *Gynecol. Oncol. Rep.*, vol. –, pp. –, Feb. 2024, doi: 10.1016/j.gore.2024.101325.

[3] S. K. Baba *et al*., "Human papilloma virus (HPV) mediated cancers: An insightful update," *J. Transl. Med.*, vol. –, pp. –, Dec. 2025, doi: 10.1186/s12967-025-06470-x.

[4] J. Wang *et al*., "Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer," *Nat. Commun.*, vol. 15, no. 1, pp. –, Dec. 2024, doi: 10.1038/s41467-024-48705-3.

[5] P. Agustiansyah *et al*., "Epidemiology and risk factors for cervical cancer," *Bioscientia Med.*, vol. 5, no. 7, pp. 624–631, Apr. 2021, doi: 10.32539/bsm.v5i7.326.

[6] A. Zumrudah *et al*., "Preferensi dan akurasi skrining kanker serviks di Indonesia: Literature review," *Sport Sci. Health*, vol. 5, no. 10, pp. 978–995, Oct. 2023, doi: 10.17977/um062v5i102023p978-995.

[7] J. H. Ma *et al*., "Computer-aided diagnosis of cervical dysplasia using colposcopic images," *Front. Oncol.*, vol.

12, pp. 905623, Aug. 2022, doi: 10.3389/fonc.2022.905623.
[8] M. M. Rahaman *et al*., "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Comput. Biol. Med.*, vol. 136, pp. 104649, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104649.
[9] F. Pesapane *et al*., "Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine," *Insights Imaging*, vol. –, pp. –, Dec. 2018, doi: 10.1186/s41747-018-0061-6.
[10] H. M. Zangana *et al*., "Exploring image representation and color spaces in computer vision: A comprehensive review," *Indones. J. Comput. Sci.*, vol. 13, no. 3, pp. –, Jun. 2024, doi: 10.33022/ijcs.v13i3.3998.
[11] A. Shiraz *et al*., "The early detection of cervical cancer: The current and changing landscape of cervical disease detection," *Cytopathology*, vol. –, pp. –, Jul. 2020, doi: 10.1111/cyt.12835.
[12] M. A. Raza *et al*., "Advanced feature extraction for cervical cancer image classification: Integrating neural feature extraction and AutoInt models," *Sensors*, vol. 25, no. 9, pp. –, May 2025, doi: 10.3390/s25092826.
[13] Y. Fan and W. Wang, "Using multi-layer perceptron to identify origins of replication in eukaryotes via informative features," *BMC Bioinf.*, vol. 22, no. 1, pp. –, Dec. 2021, doi: 10.1186/s12859-021-04431-x.
[14] C. A. Suárez *et al*., "Improving SVM performance through data reduction and misclassification analysis with linear programming," *Complex Intell. Syst.*, vol. 11, no. 8, pp. –, Aug. 2025, doi: 10.1007/s40747-025-01989-4.
[15] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. –, 2022, doi: 10.1109/ACCESS.2022.3207287.
[16] International Agency for Research on Cancer (IARC), "IARC Image Bank – Colposcopy images dataset," Dataset pribadi, 2025.
[17] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. –, Dec. 2019, doi: 10.1186/s40537-019-0192-5.
[18] F. Pedregosa *et al*., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: www.scikit-learn.sourceforge.net.