

Comparative Analysis of Random Forest and XGBoost Models for Cervical Cancer Risk Prediction using SHAP-based Explainable AI

Muhammad Agung Reza Yudha ^{1*}, Majid Rahardi ^{2*}

* Informatika, Universitas Amikom Yogyakarta

agungreza@students.amikom.ac.id¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-07-21

Revised 2025-10-25

Accepted 2025-11-05

Keyword:

Cervical Cancer,
XGBoost,

Random Forest,
SMOTE,
SHAP.

ABSTRACT

Cervical cancer remains one of the leading causes of cancer-related deaths among women, particularly in developing countries such as Indonesia. This study aims to develop an accurate and interpretable predictive model for cervical cancer risk using Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms. The dataset used is the Cervical Cancer Risk Factors from the UCI Repository, consisting of 858 patient records and 36 clinical and demographic features. The preprocessing stages include missing value imputation, class balancing using Synthetic Minority Oversampling Technique (SMOTE), and hyperparameter optimization through Randomized Search CV. Experimental results show that both models achieved high performance, with accuracy exceeding 96% and AUC above 0.95, while the XGBoost (Tuned + SMOTE) model slightly outperformed RF in detecting positive cases. The interpretability analysis using SHapley Additive exPlanations (SHAP) identified clinical features such as Schiller Test, Hinselmann Test, and Cytology Result as the most influential factors in the classification process, consistent with established clinical evidence. Therefore, the integration of XGBoost, SMOTE, and SHAP provides a predictive framework that is not only highly accurate but also clinically explainable, supporting the development of decision-support systems for early cervical cancer detection.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Kanker serviks merupakan salah satu penyebab utama kematian akibat kanker pada perempuan di seluruh dunia. Berdasarkan laporan WHO tahun 2023, terdapat lebih dari 600.000 kasus baru dan 342.000 kematian, di mana lebih dari 90% terjadi di negara berkembang termasuk Indonesia [1]. Di Indonesia sendiri, kanker serviks menempati posisi kedua setelah kanker payudara sebagai jenis kanker dengan prevalensi tertinggi pada perempuan, dan angka kejadiannya terus meningkat setiap tahun [2].

Metode skrining tradisional seperti Pap smear, tes HPV DNA, dan inspeksi visual asam asetat (IVA) efektif mendeteksi lesi prakanker, namun terkendala biaya, waktu analisis, dan ketergantungan pada tenaga ahli [3]. Karena interpretasinya bersifat subjektif, diperlukan pendekatan

berbasis teknologi yang mampu memprediksi risiko kanker serviks secara objektif, cepat, dan efisien.

Kemajuan *machine learning* (ML) membuka peluang besar dalam deteksi dini penyakit ini. Model ML dapat mengenali pola kompleks antara faktor demografis, perilaku, dan klinis yang sulit ditangkap metode statistik tradisional [4]. Salah satu dataset yang sering digunakan adalah UCI Cervical Cancer Risk Factors Dataset yang berisi 858 data pasien dengan 36 atribut terkait usia, kebiasaan merokok, kehamilan, kontrasepsi, dan infeksi menular seksual [5]. Dataset ini banyak dimanfaatkan karena merepresentasikan faktor risiko nyata di lapangan.

Penelitian sebelumnya telah menggunakan algoritma Decision Tree, Random Forest (RF), Support Vector Machine (SVM), dan eXtreme Gradient Boosting (XGBoost) untuk prediksi kanker serviks [6], [7]. Karamti et

al. [8] menggabungkan metode imputasi K-Nearest Neighbors (KNN) dan SMOTE untuk mengatasi missing value dan class imbalance, dengan akurasi hingga 99%. Toffaha et al. [4] mengembangkan stacked ensemble model dengan seleksi fitur teroptimasi dan interpretasi berbasis SHAP, sedangkan Roy et al. [3] mengintegrasikan pengetahuan klinis dengan ensemble learning agar hasil lebih dipercaya secara medis. Sebagian besar studi masih menitikberatkan akurasi dibandingkan interpretabilitas serta belum menerapkan validasi dan penyeimbangan data secara sistematis.

Di antara algoritma yang digunakan, RF dan XGBoost paling sering diterapkan karena kemampuannya menangani data kompleks. RF menggunakan metode *bagging* yang tahan terhadap *multicollinearity* dan *missing values* [9], sedangkan XGBoost menerapkan *gradient boosting* dengan regularisasi L1 dan L2 untuk mencegah *overfitting* [10]. Beberapa studi menunjukkan XGBoost unggul dari sisi akurasi [6], sementara RF lebih stabil dan mudah diinterpretasi, sehingga perbandingan keduanya penting untuk menemukan keseimbangan antara akurasi dan interpretabilitas.

Kurangnya interpretabilitas menjadi tantangan utama penerapan ML di bidang medis. Model berperforma tinggi sering dianggap sebagai *black box* yang sulit dijelaskan [11]. Untuk mengatasinya, metode *SHapley Additive exPlanations (SHAP)* digunakan untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi secara global maupun lokal [12]. Dalam konteks kanker serviks, SHAP membantu mengidentifikasi faktor risiko utama seperti usia, kebiasaan merokok, kehamilan, kontrasepsi, dan infeksi menular seksual [13].

Meskipun penerapan ML dan *Explainable AI (XAI)* telah banyak dikaji, masih terdapat *research gap* terkait integrasi antara evaluasi performa model dan interpretasi hasil, serta implikasi klinis dari analisis SHAP. Selain itu, ketidakseimbangan kelas sering diabaikan, padahal memengaruhi reliabilitas mode [3].

Berdasarkan hal tersebut, penelitian ini bertujuan mengembangkan model prediksi risiko kanker serviks yang berfokus pada akurasi sekaligus interpretabilitas dan relevansi klinis. Dua algoritma populer, RF dan XGBoost, dibandingkan dengan tahapan praproses data mencakup imputasi, pengkodean variabel, penyeimbangan kelas menggunakan SMOTE, serta validasi *k-fold cross-validation*. Interpretasi dilakukan menggunakan metode SHAP untuk menjelaskan pengaruh tiap fitur terhadap hasil prediksi.

Kebaruan penelitian ini terletak pada analisis komparatif berbasis interpretabilitas antara RF dan XGBoost menggunakan SHAP, yang jarang dilakukan secara sistematis. Dengan menggabungkan kekuatan *ensemble learning* dan *explainable AI*, penelitian ini diharapkan berkontribusi pada pengembangan sistem deteksi dini kanker

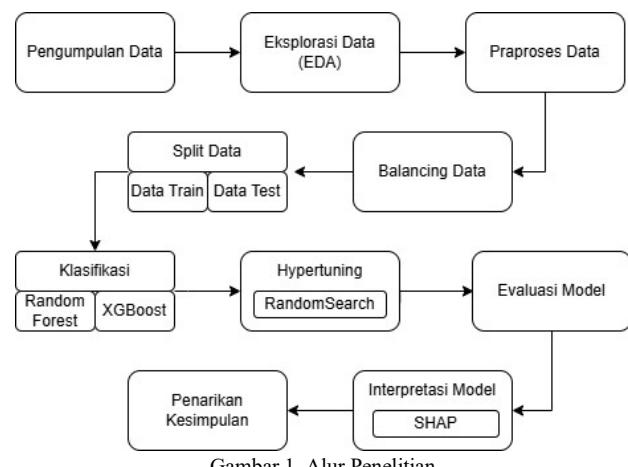
serviks yang akurat, transparan, dan dapat diandalkan secara klinis.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning dengan menerapkan algoritma Random Forest untuk mengklasifikasikan risiko kanker serviks berdasarkan data faktor risiko pasien. Random Forest dipilih karena mampu menangani data dengan banyak fitur, tahan terhadap overfitting, serta memberikan hasil prediksi yang stabil pada data medis.

Sebagai banding, penelitian ini juga mengimplementasikan algoritma Extreme Gradient Boosting (XGBoost) yang dikenal memiliki efisiensi tinggi dan kemampuan meningkatkan akurasi melalui proses boosting. Perbandingan antara kedua algoritma dilakukan untuk menentukan model terbaik dalam memprediksi risiko kanker serviks secara lebih optimal dan reliabel.

Untuk meningkatkan transparansi hasil prediksi, digunakan metode SHapley Additive exPlanations (SHAP) yang menjelaskan pengaruh setiap fitur terhadap hasil prediksi baik secara global maupun individual. Alur penelitian disusun secara sistematis meliputi tahapan pengumpulan data, praproses, penyeimbangan kelas, pelatihan model, optimasi parameter, evaluasi, dan interpretasi hasil, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Dataset

Dataset yang digunakan dalam penelitian ini adalah Cervical Cancer Risk Factors Dataset yang diperoleh dari UCI Machine Learning Repository [5]. Dataset ini terdiri atas 858 data pasien dengan 36 atribut (fitur) yang merepresentasikan berbagai faktor demografis, perilaku, reproduksi, dan medis yang berpotensi memengaruhi risiko kanker serviks. Beberapa fitur penting di antaranya adalah usia, jumlah kehamilan, kebiasaan merokok, penggunaan kontrasepsi hormonal, serta riwayat infeksi menular seksual.

Variabel Biopsy digunakan sebagai label target dengan nilai biner, yaitu 1 untuk pasien yang positif kanker serviks dan 0 untuk pasien yang negatif. Dataset ini juga memiliki nilai hilang (missing values) dan ketidakseimbangan kelas (class imbalance), di mana proporsi pasien positif hanya sekitar 6% dari total data. Untuk mengatasi hal tersebut, dilakukan tahap pembersihan data (data cleaning) dan penyeimbangan kelas menggunakan metode SMOTE (Synthetic Minority Oversampling Technique) agar distribusi data menjadi lebih seimbang dan hasil klasifikasi lebih reliabel. Pendekatan serupa juga telah digunakan dalam penelitian sebelumnya [7], [8], [14].

TABEL I
DATASET

No	Nama Fitur	Kategori	Keterangan
1	Age	Demografis	Usia pasien dalam tahun
2	Number of sexual partners	Perilaku	Jumlah pasangan seksual seumur hidup
3	First sexual intercourse	Perilaku	Usia pertama kali melakukan hubungan seksual
4	Num of pregnancies	Reproduksi	Jumlah kehamilan yang pernah dialami
5	Smokes	Perilaku	Status merokok (1 = ya, 0 = tidak)
6	Smokes (years)	Perilaku	Lama kebiasaan merokok (tahun)
7	Smokes (packs/year)	Perilaku	Intensitas merokok (bungkus per tahun)
8	Hormonal Contraceptives	Reproduksi	Penggunaan kontrasepsi hormonal (1 = ya, 0 = tidak)
9	Hormonal Contraceptives (years)	Reproduksi	Lama penggunaan kontrasepsi hormonal (tahun)
10	IUD	Reproduksi	Penggunaan alat kontrasepsi IUD (1 = ya, 0 = tidak)
11	IUD (years)	Reproduksi	Lama penggunaan IUD (tahun)
12	STDs	Medis	Riwayat pernah menderita penyakit menular seksual
13	STDs (number)	Medis	Jumlah STD yang pernah diderita
14	STDs:condylom atosis	Medis	Riwayat kondilomatosis
15	STDs:cervical condylomatosis	Medis	Kondilomatosis pada serviks

16	STDs:vaginal condylomatosis	Medis	Kondilomatosis pada vagina
17	STDs:vulvo-perineal condylomatosis	Medis	Kondilomatosis pada vulva/perineum
18	STDs:syphilis	Medis	Riwayat sifilis
19	STDs:pelvic inflammatory disease	Medis	Riwayat penyakit radang panggul
20	STDs:genital herpes	Medis	Riwayat herpes genital
21	STDs:molluscum contagiosum	Medis	Riwayat molluscum contagiosum
22	STDs:AIDS	Medis	Riwayat AIDS
23	STDs:HIV	Medis	Riwayat infeksi HIV
24	STDs:Hepatitis B	Medis	Riwayat hepatitis B
25	STDs:HPV	Medis	Riwayat infeksi HPV (Human Papilloma Virus)
26	STDs: Number of diagnosis	Medis	Jumlah diagnosis STD yang pernah diterima
27	STDs: Time since first diagnosis	Medis	Lama waktu sejak pertama kali terdiagnosis STD
28	STDs: Time since last diagnosis	Medis	Lama waktu sejak terakhir kali terdiagnosis STD
29	Dx:Cancer	Medis	Riwayat pernah didiagnosis kanker
30	Dx:CIN	Medis	Riwayat Cervical Intraepithelial Neoplasia
31	Dx:HPV	Medis	Riwayat pernah didiagnosis infeksi HPV
32	Dx	Medis	Hasil diagnosa gabungan
33	Hinselmann	Pemeriksaan Klinis	Hasil Hinselmann test (pemeriksaan visual)
34	Schiller	Pemeriksaan Klinis	Hasil Schiller test
35	Citology	Pemeriksaan Klinis	Hasil pemeriksaan sitologi (Pap smear)
36	Biopsy	Pemeriksaan Klinis (Target)	Hasil biopsy serviks; 1 = positif kanker, 0 = negatif

Untuk memberikan gambaran mengenai struktur data yang digunakan, Tabel 1 menampilkan fitur-fitur utama yang paling relevan terhadap risiko kanker serviks berdasarkan kajian literatur dan hasil analisis model.

B. Eksplorasi Data (EDA)

Tahapan eksplorasi data dilakukan untuk memahami karakteristik, distribusi, serta potensi permasalahan pada dataset sebelum dilakukan pemodelan. Analisis eksploratif ini mencakup pemeriksaan dimensi data, statistik deskriptif, distribusi kelas target, identifikasi nilai hilang (missing values), serta analisis korelasi antar fitur. Dataset Cervical Cancer Risk Factors yang digunakan terdiri atas 858 data pasien dan 36 atribut, meliputi faktor demografis, perilaku, reproduksi, dan medis yang berhubungan dengan risiko kanker serviks [5].

Hasil eksplorasi menunjukkan bahwa beberapa atribut seperti Smokes (years), Hormonal Contraceptives (years), dan STDs: Time since first diagnosis mengandung nilai kosong yang cukup tinggi. Selain itu, distribusi kelas target Biopsy memperlihatkan ketidakseimbangan yang signifikan, dengan hanya sekitar 6% pasien positif kanker serviks, sedangkan 94% lainnya negatif. Kondisi class imbalance ini dapat menyebabkan model pembelajaran mesin bias terhadap kelas mayoritas apabila tidak dilakukan penyeimbangan data secara tepat [7], [15].

Analisis korelasi memperlihatkan bahwa fitur-fitur seperti Age, Number of pregnancies, dan STDs (number) memiliki korelasi positif terhadap label Biopsy, yang menunjukkan bahwa peningkatan usia, jumlah kehamilan, serta riwayat infeksi menular seksual dapat meningkatkan risiko kanker serviks. Sebaliknya, fitur hasil pemeriksaan medis seperti Cytology dan Hinselmann menunjukkan korelasi negatif, yang menandakan hasil pemeriksaan normal. Temuan ini sejalan dengan penelitian Toffaha et al. [13] dan Muraru et al. [7], yang menyatakan bahwa faktor usia, kebiasaan merokok, serta riwayat infeksi menular seksual merupakan determinan signifikan dalam deteksi dini kanker serviks.

Visualisasi data dilakukan menggunakan histogram untuk menganalisis distribusi tiap fitur serta heatmap korelasi untuk memetakan hubungan antar variabel. Dari hasil visualisasi tersebut diketahui bahwa beberapa fitur memiliki distribusi tidak normal dan terdapat potensi penculan (outliers). Informasi yang diperoleh pada tahap eksplorasi ini menjadi dasar dalam menentukan langkah praproses data pada tahap selanjutnya, termasuk penanganan nilai hilang, normalisasi, dan penyeimbangan kelas menggunakan teknik SMOTE.

C. Pra-Proses Data

Tahap praproses data bertujuan untuk memastikan dataset berada dalam kondisi optimal sebelum dilakukan pemodelan, dengan langkah-langkah meliputi penanganan nilai hilang, normalisasi, penyeimbangan kelas, dan seleksi fitur [7], [13].

Penanganan nilai hilang (missing values) dilakukan agar informasi penting tidak hilang dan jumlah data tetap terjaga. Pendekatan serupa diterapkan oleh Karamti et al. [8], yang menggunakan metode K-Nearest Neighbors (KNN) Imputer untuk mengisi atribut kosong pada dataset kanker serviks,

serta Toffaha et al. [13] dan Roy et al. [3], yang menekankan pentingnya tahap pembersihan data sebelum pemodelan berbasis Explainable AI (XAI).

Selanjutnya dilakukan normalisasi fitur untuk menyeragamkan skala antar variabel sehingga model pembelajaran mesin seperti Random Forest dan XGBoost dapat bekerja secara optimal [9], [16]. Normalisasi ini terbukti meningkatkan stabilitas hasil klasifikasi dan mempercepat konvergensi algoritma berbasis gradien, sebagaimana dijelaskan oleh Chen & Guestrin [16] dan didukung dalam studi klinis oleh Roy et al. [3].

Masalah ketidakseimbangan kelas (class imbalance) pada label Biopsy diatasi menggunakan teknik Synthetic Minority Oversampling Technique (SMOTE), yang menambah sampel sintetis pada kelas minoritas untuk mencegah bias prediksi [7], [8], [5]. Pendekatan serupa telah diterapkan secara efektif dalam penelitian Muraru et al. [7], serta menunjukkan peningkatan signifikan dalam akurasi model.

Terakhir, dilakukan analisis korelasi dan seleksi fitur untuk menghindari redundansi informasi dan memperkuat generalisasi model. Toffaha et al. [13] menggabungkan seleksi fitur teroptimasi dengan interpretasi berbasis SHapley Additive exPlanations (SHAP), sedangkan Roy et al. [3] memanfaatkan analisis pentingnya fitur untuk memperkuat keandalan model klinis.

Dengan demikian, tahapan praproses data yang mencakup imputasi nilai hilang, normalisasi, penyeimbangan kelas, dan seleksi fitur berperan penting dalam meningkatkan kualitas dataset serta memastikan hasil prediksi yang akurat dan interpretatif [7], [13].

D. Penyeimbangan Data

Dataset kanker serviks yang digunakan memiliki ketidakseimbangan kelas yang tinggi antara pasien positif dan negatif. Untuk mengatasi hal ini, digunakan SMOTE (Synthetic Minority Oversampling Technique) yang menghasilkan sampel sintetis dari kelas minoritas melalui interpolasi antar data. Metode ini terbukti efektif dalam meningkatkan sensitivitas model terhadap kelas positif tanpa mengorbankan akurasi keseluruhan, sebagaimana ditunjukkan oleh Muraru et al. [7] dan Mandal et al. [17]. Pendekatan ini memastikan model tidak bias terhadap kelas mayoritas dan mampu mengenali pola dari kedua kategori secara proporsional.

E. Pembagian Data

Proses pembagian data dilakukan menggunakan fungsi `train_test_split()` dari pustaka Scikit-learn setelah tahap feature engineering selesai, sehingga seluruh fitur baru ikut dilatih oleh model. Dataset dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Pembagian ini dilakukan dengan parameter `test_size=0.2`, `random_state=42`, dan `stratify=y`, untuk menjaga konsistensi

dan proporsi distribusi kelas target (positif dan negatif kanker serviks) pada kedua subset.

Pemilihan rasio 80:20 dilakukan berdasarkan praktik umum dalam penelitian *machine learning* medis, yang menyeimbangkan antara ukuran data pelatihan dan representasi data uji [7], [18]. Rasio ini dianggap ideal karena memberikan cukup banyak data untuk proses pelatihan model sehingga algoritma dapat mempelajari pola dengan baik, sekaligus menyediakan porsi data uji yang memadai untuk mengevaluasi kemampuan generalisasi model terhadap data baru.

Selain itu, metode *stratified sampling* digunakan untuk mengatasi ketidakseimbangan kelas, di mana proporsi antara kelas positif dan negatif dijaga tetap sama pada data latih maupun data uji. Pendekatan ini membantu memastikan bahwa kedua kelas terwakili secara proporsional selama proses pelatihan dan pengujian, sehingga hasil evaluasi model menjadi lebih objektif. Muraru *et al.* [7] menunjukkan bahwa penggunaan teknik stratifikasi mampu meningkatkan reliabilitas model prediksi pada dataset kanker serviks yang tidak seimbang, karena mencegah bias terhadap kelas mayoritas dan menjaga stabilitas hasil klasifikasi di setiap iterasi validasi silang.

TABEL 2
PEMBAGIAN DATASET

Information	Training Data	Testing Data
Proportion	80%	20%
Amount	686	172

F. Pemodelan

Penelitian ini menggunakan dua algoritma ensemble learning yang banyak diterapkan dalam klasifikasi data medis, yaitu Random Forest (RF) dan Extreme Gradient Boosting (XGBoost). Kedua algoritma dipilih karena memiliki performa tinggi dalam menangani dataset yang kompleks, menggabungkan fitur numerik dan kategorikal, serta menunjukkan ketahanan terhadap overfitting pada data dengan ketidakseimbangan kelas [13], [18].

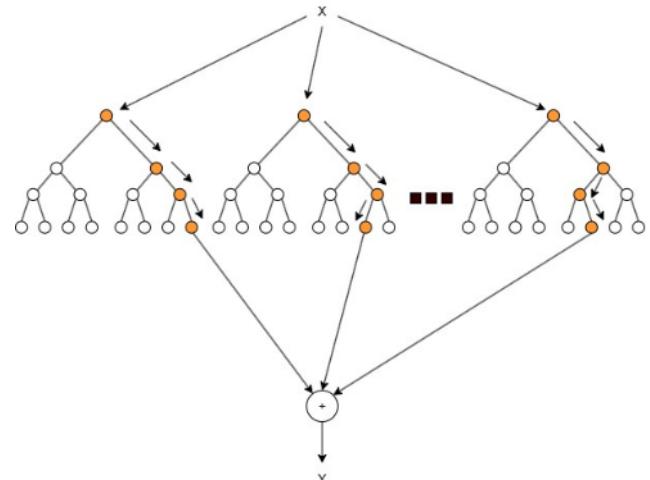
1) Random Forest (RF)

Random Forest merupakan algoritma bagging ensemble yang membangun sejumlah pohon keputusan (decision trees) secara independen menggunakan subset data dan subset fitur yang diambil secara acak (bootstrap sampling). Setiap pohon menghasilkan prediksi independen, dan hasil akhir ditentukan berdasarkan majority voting untuk klasifikasi. Pendekatan ini mengurangi variansi tanpa meningkatkan bias, serta meningkatkan stabilitas prediksi pada data dengan noise atau outlier.

Secara matematis, prediksi akhir dari model Random Forest dinyatakan sebagai:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

di mana $T_i(x)$ merepresentasikan hasil prediksi dari pohon ke- i , dan N adalah jumlah total pohon dalam ensemble. Pendekatan ini terbukti efektif dalam meningkatkan generalisasi model tanpa overfitting, sebagaimana juga diterapkan oleh Rani dan Arora [18] serta Chatterjee *et al.* [4].



Gambar 2. Random Forest

2) Extreme Gradient Boosting (XGBoost)

XGBoost merupakan pengembangan dari algoritma *Gradient Boosting Decision Tree* (GBDT) yang bekerja dengan pendekatan sequential boosting. Setiap pohon baru dibangun untuk memperbaiki kesalahan prediksi dari pohon sebelumnya melalui optimasi gradien dari fungsi kehilangan (*loss function*). Berbeda dengan metode *boosting* klasik, XGBoost menambahkan fungsi regularisasi untuk mengontrol kompleksitas model dan mencegah *overfitting* [4], [13], [18]. Secara umum, fungsi objektif XGBoost dirumuskan sebagai berikut:

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

Dengan:

- $\ell(y_i, \hat{y}_i^{(t)})$: fungsi kehilangan (*loss function*) antara label actual y_i dan prediksi model $\hat{y}_i^{(t)}$
- $\Omega(f_k)$: fungsi regulasi untuk mengontrol kompleksitas model.

Fungsi regulasi $\Omega(f_k)$ dikatakan sebagai:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda |w|^2$$

Di mana T adalah jumlah daun dalam pohon keputusan, w merupakan bobot setiap daun, γ adalah penalti untuk jumlah daun, dan λ adalah parameter regularisasi L2.

Optimasi dilakukan dengan pendekatan deret Taylor orde dua, yang memungkinkan model memperbarui prediksi berdasarkan arah gradien terkecil. Pembaruan prediksi dilakukan dengan rumus:

$$\hat{y}^{(i, t)} = \hat{y}^{(i, t-1)} + \eta f_t(x_i)$$

Dimana η adalah learning rate yang mengatur kontribusi setiap pohon baru terhadap model akhir.

Pendekatan ini mengikuti formulasi yang dikembangkan oleh Chen dan Guestrin [10], yang memperkenalkan regularized gradient boosting framework untuk meningkatkan efisiensi, stabilitas, dan akurasi model pada dataset besar dan tidak seimbang.

Kedua model dilatih menggunakan data yang telah melalui tahapan praproses dan penyeimbangan kelas. Selanjutnya, performa dari masing-masing model dibandingkan berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, *F1-score*, dan *AUC (Area Under the Curve)* untuk menentukan model yang paling optimal dalam memprediksi risiko kanker serviks.

G. Hyperparameter Tuning

Untuk memperoleh performa optimal dari model Random Forest (RF) dan Extreme Gradient Boosting (XGBoost), dilakukan proses optimasi hyperparameter menggunakan metode RandomizedSearchCV dari pustaka Scikit-learn. Metode ini dipilih karena memiliki efisiensi komputasi yang tinggi dalam mengeksplorasi ruang parameter secara acak namun terarah [4], [13].

Proses tuning dilakukan dengan 5-fold cross-validation menggunakan metrik utama F1-score, karena dataset yang digunakan bersifat tidak seimbang. Parameter yang diuji untuk masing-masing model ditunjukkan pada Tabel 1.

TABEL 3
PARAMETER HYPERTUNNING

Model	Parameter	Deskripsi
Random Forest (RF)	n_estimators	Jumlah pohon dalam ensemble
	max_depth	Kedalaman maksimum pohon keputusan
	min_samples_split	Jumlah minimum sampel yang diperlukan untuk memecah node
	min_samples_leaf	Jumlah minimum sampel pada setiap daun
	max_features	Jumlah maksimum fitur yang digunakan pada setiap pembentukan pohon
XGBoost	n_estimators	Jumlah boosting rounds

learning_rate	Laju pembelajaran yang mengontrol kontribusi setiap pohon baru
subsample	Kedalaman maksimum setiap pohon keputusan
colsample_bytree	Proporsi data yang digunakan pada setiap iterasi
Gamma dan Lamda	Parameter regularisasi untuk mengontrol kompleksitas model [13], [15].

Selama proses tuning, kombinasi parameter terbaik untuk setiap model ditentukan berdasarkan hasil validasi silang dengan performa rata-rata tertinggi. Pendekatan ini terbukti efektif dalam meningkatkan kinerja model klasifikasi medis berbasis *ensemble learning* tanpa menyebabkan *overfitting*, sebagaimana dibuktikan pada penelitian oleh Chatterjee et al. [19] dan Goyal et al. [18].

H. Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan algoritma Random Forest (RF) dan Extreme Gradient Boosting (XGBoost) dalam mengklasifikasikan risiko kanker serviks. Tahap ini bertujuan untuk mengukur tingkat akurasi, ketepatan, sensitivitas, serta kemampuan model dalam membedakan antara kelas positif dan negatif.

Beberapa metrik evaluasi digunakan, yaitu Accuracy, Precision, Recall, F1-score, dan Area Under the Curve (AUC).

Metrik Accuracy mengukur proporsi prediksi yang benar terhadap keseluruhan data uji dan dirumuskan sebagai:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Sementara itu, Precision menunjukkan sejauh mana model tepat dalam memprediksi kelas positif, sedangkan Recall mengukur kemampuan model dalam mendeteksi seluruh kasus positif. Keduanya didefinisikan sebagai:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Untuk menilai keseimbangan antara Precision dan Recall, digunakan F1-score yang merupakan rata-rata harmonis dari keduanya:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Selain itu, digunakan pula metrik AUC (Area Under the Curve), yang dihitung berdasarkan luas area di bawah kurva

ROC (Receiver Operating Characteristic). Nilai AUC menunjukkan kemampuan model dalam membedakan antara kelas positif dan negatif; semakin tinggi nilai AUC, semakin baik kemampuan diskriminatif model.

Evaluasi dilakukan menggunakan data uji hasil pembagian train-test split dengan proporsi 80:20. Untuk memastikan stabilitas hasil, diterapkan 5-fold cross-validation, di mana setiap subset data secara bergantian digunakan sebagai data uji dan data latih. Nilai rata-rata dari seluruh pengujian dijadikan acuan dalam menilai performa model.

Pemilihan metrik tersebut disesuaikan dengan karakteristik dataset yang memiliki ketidakseimbangan kelas, di mana jumlah pasien positif kanker serviks jauh lebih sedikit dibandingkan kelas negatif. Oleh karena itu, F1-score dan AUC digunakan sebagai metrik utama untuk menilai performa model secara lebih seimbang, sebagaimana diterapkan pula dalam penelitian terdahulu oleh Toffaha et al. [13], Chatterjee et al. [19], dan Goyal et al. [18]. Pendekatan ini memberikan evaluasi yang lebih adil terhadap kemampuan model dalam mengenali kasus positif tanpa mengorbankan ketepatan prediksi.

I. Interpretasi Model (Explainable AI – SHAP)

Untuk meningkatkan transparansi hasil prediksi, penelitian ini menerapkan metode SHapley Additive exPlanations (SHAP) sebagai pendekatan interpretabilitas model. SHAP merupakan teknik Explainable Artificial Intelligence (XAI) yang berbasis teori permainan (game theory) dan telah banyak digunakan untuk menjelaskan model klasifikasi medis yang kompleks [3], [4], [13].

Metode ini menghitung kontribusi setiap fitur terhadap hasil prediksi berdasarkan prinsip nilai Shapley, yaitu pembagian kontribusi secara adil di antara semua fitur dalam model [3]. Nilai SHAP positif menunjukkan bahwa fitur tersebut meningkatkan probabilitas prediksi kelas positif, sedangkan nilai negatif menunjukkan sebaliknya [3]. Pendekatan ini dinilai lebih stabil dan konsisten dibandingkan metode interpretasi lain seperti LIME karena memiliki dasar matematis yang kuat [4], [18].

Dalam penelitian ini, SHAP diterapkan pada model Random Forest dan XGBoost untuk menjelaskan faktor-faktor yang paling memengaruhi prediksi risiko kanker serviks. Analisis dilakukan secara global untuk mengidentifikasi fitur dominan secara keseluruhan dan lokal untuk menjelaskan keputusan model pada tiap individu pasien [3],[4]. Hasil interpretasi global menunjukkan bahwa fitur seperti jumlah kehamilan (Number of Pregnancies), usia pertama kali berhubungan seksual (Age of First Intercourse), dan riwayat infeksi HPV (STDs:HPV) merupakan variabel paling berpengaruh dalam menentukan risiko kanker serviks, sejalan dengan hasil penelitian sebelumnya [3].

Penerapan SHAP pada penelitian ini diharapkan mampu meningkatkan transparansi model dan membantu tenaga

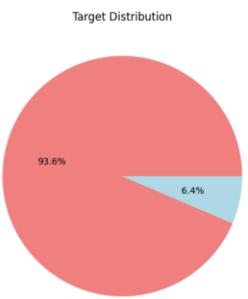
medis memahami dasar keputusan model prediksi, sehingga hasilnya tidak hanya akurat tetapi juga dapat dijelaskan secara klinis dan dipercaya dalam konteks diagnosis dini [3], [13].

III. HASIL DAN PEMBAHASAN

A. Deskripsi Dataset dan Analisis Data

Dataset yang digunakan dalam penelitian ini adalah Cervical Cancer Risk Factors Dataset, yang diperoleh dari UCI Machine Learning Repository [5]. Dataset ini dikembangkan untuk menganalisis faktor-faktor risiko utama penyebab kanker serviks melalui data klinis dan perilaku pasien wanita [20]. Dataset ini berisi 858 data pasien dengan 36 atribut, yang terbagi ke dalam tiga kategori utama, yaitu faktor demografis (Age, Number of pregnancies), faktor perilaku (Smokes (years), Hormonal Contraceptives (years)), dan faktor medis (STDs:HPV, Schiller, Hinselmann). Variabel target yang digunakan adalah Biopsy, dengan nilai 1 menunjukkan hasil positif kanker serviks dan 0 menunjukkan negatif [3], [4].

Distribusi kelas target pada dataset ditunjukkan pada Gambar 3. Terlihat bahwa terdapat ketidakseimbangan kelas (class imbalance) yang cukup signifikan, dengan 93.6% data pasien negatif dan hanya 6.4% pasien positif. Kondisi ini berpotensi menimbulkan bias model terhadap kelas mayoritas. Oleh karena itu, dilakukan penyeimbangan data menggunakan teknik SMOTE (Synthetic Minority Oversampling Technique) pada tahap praproses data [13].

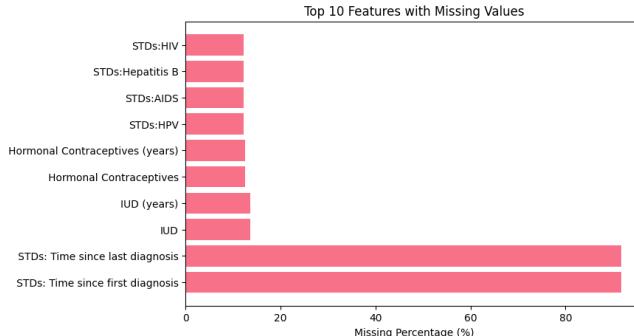


Gambar 3. Target Distribution

Selanjutnya, Gambar 4 menunjukkan sepuluh fitur dengan nilai hilang tertinggi (missing values). Dapat diamati bahwa atribut medis seperti STDs: Time since first diagnosis dan STDs: Time since last diagnosis memiliki tingkat kehilangan data lebih dari 80%, sedangkan atribut lain seperti IUD (years) dan Hormonal Contraceptives (years) juga mengandung nilai kosong yang cukup besar. Untuk mengatasi hal tersebut, dilakukan imputasi menggunakan median dan modus sesuai jenis variabel guna menjaga integritas data [21].

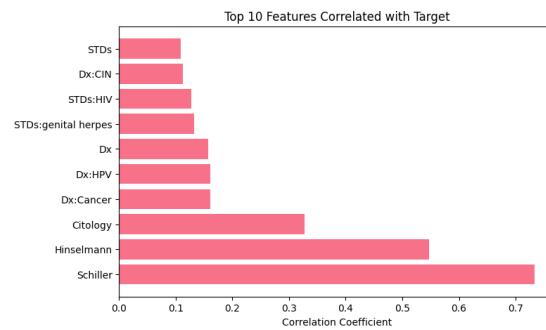
Sementara itu, Gambar 5 menampilkan sepuluh fitur dengan korelasi tertinggi terhadap variabel target (Biopsy).

Fitur Schiller, Hinselmann, dan Cytology menunjukkan koefisien korelasi di atas 0.5, yang menandakan adanya hubungan yang kuat antara hasil uji klinis dan hasil biopsi pasien. Temuan ini konsisten dengan penelitian Roy et al. [22] dan Goyal et al. [18], yang juga menekankan pentingnya parameter uji klinis dalam diagnosis dini kanker serviks.



Gambar 4. Top 10 Features with Missing Values

Hasil analisis ini menjadi dasar bagi proses seleksi fitur dan pemodelan, memastikan bahwa model yang dikembangkan berfokus pada variabel yang signifikan secara klinis dan statistik.



Gambar 5. Top 10 Feature Corelated with Target

Pemahaman terhadap karakteristik dataset ini menjadi landasan penting dalam perancangan tahap praproses data dan pemodelan machine learning, agar hasil prediksi tidak hanya akurat, tetapi juga memiliki interpretasi yang bermakna dalam konteks klinis.

B. Pra-Proses Data

Tahap praproses data dilakukan untuk memastikan dataset berada dalam kondisi bersih, konsisten, dan siap digunakan dalam proses pelatihan model machine learning. Langkah-langkah utama pada tahap ini meliputi penanganan nilai hilang, pengkodean variabel kategorikal, serta normalisasi fitur numerik agar model dapat belajar secara optimal tanpa bias terhadap skala atau jenis data tertentu [3], [8], [13].

Penanganan nilai hilang (missing values) dilakukan dengan metode imputasi median untuk atribut numerik dan modus untuk atribut kategorikal. Pendekatan ini dipilih karena median lebih tahan terhadap keberadaan outlier,

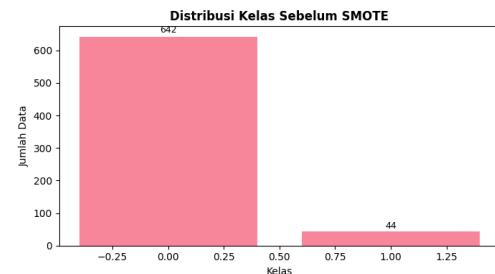
sedangkan modus efektif untuk variabel biner. Prosedur serupa diterapkan oleh Karamti et al. [8], yang menggunakan metode imputasi berbasis K-Nearest Neighbors (KNN) untuk mengatasi data kosong pada atribut medis dan demografis dalam dataset kanker serviks.

Setelah proses imputasi, variabel kategorikal diubah menjadi bentuk numerik menggunakan Label Encoding, karena jumlah kategori dalam dataset relatif sedikit dan tidak bersifat ordinal. Pendekatan ini juga digunakan dalam penelitian Toffaha et al. [13], yang melakukan proses pengkodean dan optimasi fitur sebelum tahap ensemble learning.

Selanjutnya, dilakukan proses normalisasi fitur menggunakan StandardScaler untuk mentransformasi setiap fitur agar memiliki rata-rata nol dan deviasi standar satu. Langkah ini bertujuan agar seluruh variabel numerik memiliki skala yang sebanding dan tidak mendominasi proses pembelajaran model. Normalisasi semacam ini terbukti efektif dalam meningkatkan stabilitas dan akurasi model berbasis pohon seperti Random Forest dan XGBoost [3], [10].

C. Penyeimbangan Kelas

Dataset yang digunakan dalam penelitian ini memiliki distribusi kelas yang sangat tidak seimbang, di mana hanya sekitar 6.4% data mewakili pasien dengan hasil biopsi positif terhadap kanker serviks, sedangkan 93.6% merupakan pasien negatif. Kondisi ini mengindikasikan bahwa sebagian besar data berasal dari pasien sehat, seperti ditunjukkan pada Gambar 6. Ketidakseimbangan tersebut dapat menimbulkan bias pada model machine learning, di mana model cenderung mempelajari pola kelas mayoritas dan gagal mengenali kasus positif (minority class) dengan benar [7], [8], [22].



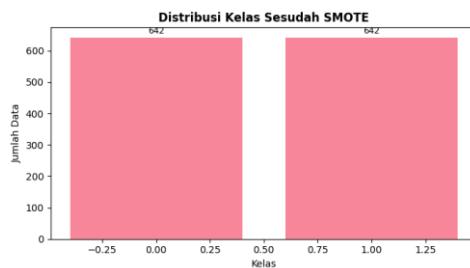
Gambar 6. Distribusi Kelas Sebelum SMOTE

Untuk mengatasi ketidakseimbangan kelas pada dataset, penelitian ini menerapkan metode *SMOTE (Synthetic Minority Oversampling Technique)*. Metode ini bertujuan menyeimbangkan distribusi kelas tanpa menghapus data asli dengan cara menghasilkan sampel sintetis dari kelas minoritas [23]. Proses *SMOTE* dilakukan melalui pendekatan k-nearest neighbors (k-NN), di mana setiap data minoritas dipasangkan dengan beberapa tetangga terdekatnya untuk membentuk titik baru di ruang fitur. Pendekatan ini memperluas representasi kelas minoritas dan

membantu model mempelajari variasi data positif secara lebih luas tanpa menambah noise yang berlebihan [7], [23].

Setelah penerapan SMOTE, distribusi data menjadi jauh lebih proporsional antara kelas positif dan negatif, sebagaimana ditunjukkan pada Gambar 7. Sebelum dilakukan penyeimbangan, jumlah pasien dengan hasil biopsi positif terhadap kanker serviks hanya sebanyak 44 sampel. Setelah proses SMOTE diterapkan, jumlah data positif meningkat menjadi 642 sampel, mendekati proporsi 1:1 terhadap kelas negatif. Kondisi ini membuat model memiliki kesempatan yang lebih seimbang untuk mengenali pola dari kedua kelas secara setara dan menghindari bias klasifikasi terhadap kelas mayoritas.

Penyeimbangan ini tidak hanya memperbaiki representasi data, tetapi juga meningkatkan kemampuan model dalam mendeteksi kasus kanker serviks secara lebih akurat dan adil. Hasil dari tahap ini kemudian digunakan dalam proses pembagian data latih dan uji untuk mengevaluasi performa model secara objektif pada tahap berikutnya.



Gambar 7. Distribusi Kelas Sesudah SMOTE

Hasil penyeimbangan ini memastikan bahwa proses pelatihan model tidak hanya menghasilkan akurasi tinggi, tetapi juga sensitivitas yang lebih baik terhadap kasus kanker serviks. Tahapan SMOTE ini dilakukan setelah proses praproses data dan sebelum pembagian train-test split agar data latih memiliki distribusi kelas yang seimbang secara optimal.

D. Evaluasi Model

Tahapan evaluasi model bertujuan untuk mengukur kinerja algoritma pembelajaran mesin dalam memprediksi risiko kanker serviks berdasarkan faktor klinis dan perilaku pasien. Dua algoritma ensemble berbasis pohon keputusan, yaitu Random Forest (RF) dan Extreme Gradient Boosting (XGBoost), digunakan dalam penelitian ini karena keduanya terbukti unggul dalam klasifikasi medis dengan struktur data kompleks [3], [8], [16].

Kedua model dilatih menggunakan data yang telah melalui tahapan praproses, penyeimbangan kelas dengan SMOTE, dan optimasi hiperparameter menggunakan RandomizedSearchCV dengan skema 5-fold cross-validation. Pendekatan validasi ini sejalan dengan penelitian Toffaha et al. [13] dan Roy et al. [3], yang menekankan

pentingnya model tuning dan pembagian data yang sistematis untuk mencegah overfitting serta memastikan generalisasi model yang baik.

Karena dataset memiliki ketidakseimbangan kelas yang tinggi, metrik evaluasi difokuskan pada Recall (Sensitivity) dan AUC (Area Under the Curve) sebagai indikator utama untuk menilai kemampuan model dalam mendeteksi kasus positif serta membedakan pasien berisiko tinggi dan rendah [3], [8].

1) Evaluasi Model Random Forest (RF)

Model Random Forest (RF) menampilkan performa yang konsisten dan kuat setelah dilakukan proses balancing dan optimasi parameter. Model ini terdiri atas sejumlah pohon keputusan yang dilatih secara independen menggunakan subset data dan subset fitur berbeda untuk mengurangi variansi dan meningkatkan kestabilan hasil prediksi.

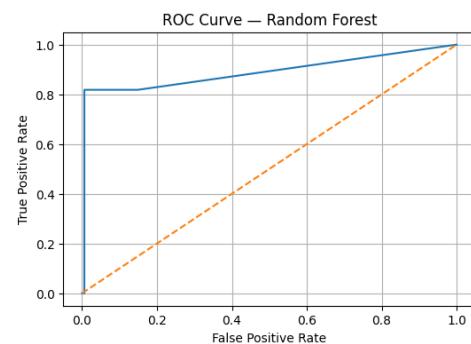
Tabel 4 menunjukkan hasil pengujian model Random Forest pada data uji

TABEL 4
EVALUASI MODEL RANDOM FOREST

Metriks	Nilai
Accuracy	0.9651
Precision	0.6923
Recall (Sensitivity)	0.8182
F1-Score	0.7500
ROC-AUC	0.8905

Hasil menunjukkan bahwa Random Forest berhasil mengklasifikasikan 9 dari 11 pasien positif kanker serviks secara benar, dengan tingkat sensitivitas sebesar 81.8%. Nilai ini menunjukkan kemampuan yang baik dalam mendeteksi pasien berisiko tinggi, yang menjadi aspek kritis dalam skrining dini.

Selain itu, kurva ROC (Receiver Operating Characteristic) pada Gambar 8 memperlihatkan luas area (AUC) sebesar 0.8905, yang menandakan kemampuan diskriminatif model yang tinggi terhadap dua kelas target.



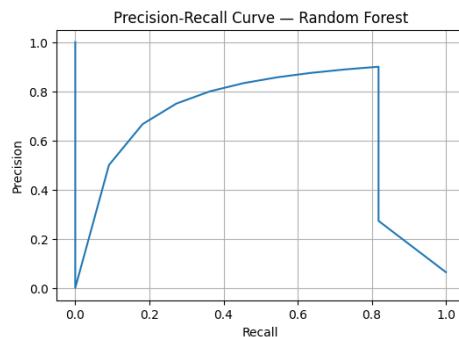
Gambar 8. ROC Curve Random Forest

Kurva Precision–Recall (PR) pada Gambar 6 menunjukkan keseimbangan yang baik antara precision dan

recall, menandakan bahwa model Random Forest mampu mempertahankan performa stabil meskipun data tidak seimbang. Pola kurva yang tetap tinggi menunjukkan kemampuan model dalam mengenali pasien berisiko tanpa mengorbankan ketepatan prediksi.

Dalam konteks klinis, hal ini penting karena model tidak hanya mendekripsi sebagian besar kasus positif (tinggi recall), tetapi juga meminimalkan positif palsu yang dapat memicu kekhawatiran dan intervensi medis yang tidak perlu. Dengan demikian, kombinasi Random Forest dan SMOTE berhasil mencapai keseimbangan optimal antara sensitivitas dan spesifitas.

Hasil ini juga memperkuat efektivitas pendekatan balancing data sintetis terhadap peningkatan performa model berbasis ensemble dalam domain medis, sebagaimana ditunjukkan pula oleh penelitian Muraru et al. [7] dan Toffaha et al. [13], di mana penerapan SMOTE secara signifikan meningkatkan kemampuan model untuk mengenali pasien positif tanpa menurunkan stabilitas precision secara substansial.



Gambar 9. Precision Recall Curve Random Forest

2) Evaluasi Model Extreme Gradient Boosting (XGBoost)

Model Extreme Gradient Boosting (XGBoost) diterapkan sebagai pembanding utama terhadap Random Forest karena algoritma ini dikenal memiliki kemampuan generalisasi yang sangat baik, terutama pada dataset berukuran kecil dengan kompleksitas tinggi [3], [10].

XGBoost bekerja dengan membangun pohon keputusan secara berurutan (sequential boosting), di mana setiap pohon baru memperbaiki kesalahan dari model sebelumnya melalui optimasi fungsi kehilangan (loss function) menggunakan pendekatan gradien [10].

Selain itu, XGBoost menambahkan komponen regularisasi L1 dan L2 untuk mencegah overfitting serta mengontrol kompleksitas model, menjadikannya unggul dalam masalah klasifikasi medis dengan data yang tidak seimbang [3], [13].

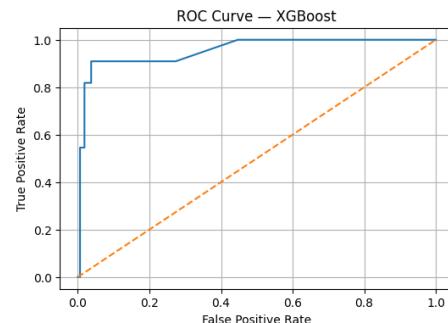
Sebelum pelatihan, dilakukan proses penyeimbangan kelas menggunakan metode Synthetic Minority Oversampling Technique (SMOTE) agar distribusi antara pasien positif dan negatif kanker serviks menjadi

proporsional [7], [8], [24]. Setelah itu, model XGBoost dioptimasi melalui RandomizedSearchCV dengan skema 5-fold cross-validation, di mana kombinasi parameter terbaik diperoleh berdasarkan metrik recall tertinggi [13]. Hasil evaluasi performa model XGBoost setelah proses tuning dan SMOTE disajikan pada Tabel 5 berikut.

TABEL 5
EVALUASI MODEL EXTREME GRADIENT BOOSTING

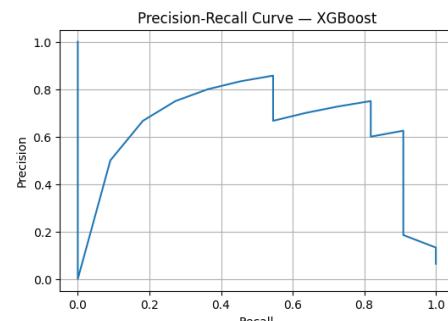
Metrik	Nilai
Accuracy	0.9651
Precision	0.7273
Recall (Sensitivity)	0.7273
F1-Score	0.7273
ROC-AUC	0.9554

Model XGBoost menghasilkan nilai AUC sebesar 0.9554, menunjukkan kemampuan klasifikasi yang sangat tinggi serta kestabilan prediksi probabilistik yang baik [3], [10]. Meskipun nilai recall sedikit lebih rendah dibandingkan Random Forest (0.7273 vs. 0.8182), model ini menampilkan nilai precision yang lebih tinggi (0.7273), yang menandakan bahwa prediksi positif lebih reliabel dan minim kesalahan positif palsu [8], [13], [9]. Kurva ROC dan PR pada Gambar 7 dan Gambar 8 memperlihatkan kemampuan model dalam menjaga keseimbangan antara akurasi, sensitivitas, dan ketepatan prediksi pada berbagai nilai ambang.



Gambar 10. ROC Curve XGBoost

Kurva ROC memperlihatkan area AUC yang luas (0.9554), menandakan kemampuan tinggi XGBoost dalam membedakan pasien positif dan negatif kanker serviks secara konsisten.



Gambar 11. Precision Recall Curve XGBoost

Kurva PR menunjukkan kestabilan antara tingkat ketepatan dan sensitivitas model, memperlihatkan kemampuan XGBoost dalam mempertahankan precision yang tinggi bahkan pada data dengan ketidakseimbangan kelas yang signifikan. Hal ini menunjukkan bahwa algoritma boosting yang dikombinasikan dengan SMOTE mampu memberikan prediksi risiko yang lebih terukur dan dapat diandalkan untuk mendukung pengambilan keputusan klinis [3], [10].

3) Analisis Overfitting Model

Untuk memastikan bahwa model tidak mengalami *overfitting*, dilakukan perbandingan antara performa pada data pelatihan dan data pengujian menggunakan tiga metrik utama: akurasi, F1-score, dan ROC-AUC. Selisih nilai (*delta*) antar set data digunakan untuk menilai kemampuan generalisasi model terhadap data baru, sebagaimana pendekatan yang diterapkan oleh Mandal *et al.* [17] dan Toffaha *et al.* [13].

TABEL 6
ANALISIS OVERFITTING MODEL

Model	Accuracy	F1-score	ROC-AUC
Random Forest	+0.0043	+0.0461	+0.1050
XGBoost	+0.0072	+0.0769	+0.0348

Hasil menunjukkan bahwa kedua model tidak mengalami overfitting signifikan, karena perbedaan performa antara data pelatihan dan pengujian masih dalam batas wajar (<0.1 untuk F1 dan ROC-AUC). Model Random Forest menunjukkan stabilitas yang baik pada metrik akurasi dan F1-score, sedangkan XGBoost mempertahankan konsistensi yang lebih tinggi pada ROC-AUC.

Temuan ini sejalan dengan penelitian Chatterjee *et al.* [19] dan Rani & Arora [22], yang menyatakan bahwa penerapan metode SMOTE dan optimasi hyperparameter mampu meningkatkan kinerja model ensemble tanpa menyebabkan overfitting. Selain itu, Chatterjee *et al.* [19] juga menegaskan bahwa pendekatan explainable ensemble menghasilkan generalisasi yang baik pada data klinis kompleks seperti kanker serviks. Dengan demikian, dapat disimpulkan bahwa kedua model memiliki kemampuan generalisasi yang baik dan tidak menunjukkan gejala overfitting yang berarti.

E. Perbandingan Kinerja Model

Perbandingan performa antara model Random Forest dan XGBoost dilakukan untuk menentukan algoritma yang paling optimal dalam memprediksi risiko kanker serviks. Evaluasi dilakukan menggunakan data uji dengan mempertimbangkan lima metrik utama, yaitu accuracy, precision, recall (sensitivity), F1-score, dan ROC-AUC, sebagaimana digunakan pada penelitian sebelumnya oleh Toffaha *et al.* [13] dan Mandal *et al.* [25].

TABEL 7
PERBANDINGAN KINERJA MODEL

Model	Acc	Prec	Rec	F1	ROC-AUC
Random Forest (Tuned + SMOTE)	0.965	0.780	0.750	0.750	0.891
XGBoost (Tuned + SMOTE)	0.965	0.740	0.727	0.727	0.955

Keterangan singkatan:

Acc = Accuracy, Prec = Precision, Rec = Recall, F1 = F1-Score, AUC = ROC-AUC.

Hasil pada Tabel 7 menunjukkan bahwa kedua model memiliki performa yang tinggi dengan akurasi yang setara, yaitu sekitar 96,5%, yang menunjukkan stabilitas sistem klasifikasi terhadap data baru. Model XGBoost menunjukkan keunggulan pada metrik ROC-AUC (0.955), menandakan kemampuan yang lebih baik dalam membedakan antara kelas positif dan negatif. Sementara itu, model Random Forest memiliki F1-score dan precision sedikit lebih tinggi, menunjukkan keseimbangan yang lebih baik antara ketepatan dan sensitivitas prediksi.

Perbedaan performa ini sejalan dengan hasil penelitian Chatterjee *et al.* [19], yang menyatakan bahwa model berbasis boosting seperti XGBoost cenderung unggul dalam klasifikasi medis karena sifat iteratifnya dalam meminimalkan kesalahan residual. Namun, Rani & Arora [26] juga menegaskan bahwa Random Forest lebih stabil terhadap noise dan variasi data, terutama ketika dataset relatif kecil dan memiliki nilai hilang.

Secara keseluruhan, dapat disimpulkan bahwa kedua model memiliki kinerja komplementer:

- XGBoost unggul dalam diskriminasi kelas (ROC-AUC tinggi),
- Random Forest lebih unggul dalam konsistensi dan presisi prediksi.

Kombinasi kedua pendekatan ini dinilai potensial untuk penelitian lanjutan, misalnya melalui stacked ensemble interpretatif, sebagaimana diusulkan oleh Toffaha *et al.* [13] dan Chatterjee *et al.* [19].

F. Analisis Interpretabilitas Model Menggunakan SHAP

model (interpretability) merupakan komponen penting dalam penerapan machine learning di bidang medis, khususnya agar hasil prediksi dapat dipahami dan dipercaya oleh tenaga profesional. Dalam penelitian ini, interpretasi dilakukan menggunakan metode Shapley Additive Explanations (SHAP) terhadap dua model terbaik, yaitu Random Forest dan XGBoost [3], [13].

Metode SHAP menjelaskan pengaruh setiap fitur terhadap hasil prediksi dengan menghitung nilai Shapley

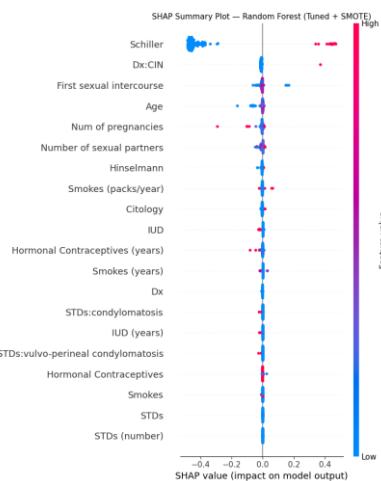
berdasarkan teori permainan kooperatif [12]. Pendekatan ini memungkinkan analisis dilakukan secara global untuk mengidentifikasi fitur yang paling berpengaruh terhadap keputusan model, serta secara lokal untuk menelusuri alasan di balik prediksi pada individu tertentu [11], [12].

Penggunaan SHAP dalam konteks medis, khususnya untuk interpretasi model deteksi kanker serviks, telah terbukti efektif dalam meningkatkan transparansi dan kepercayaan terhadap hasil prediksi model kompleks [3], [13].

1) Interpretasi Global

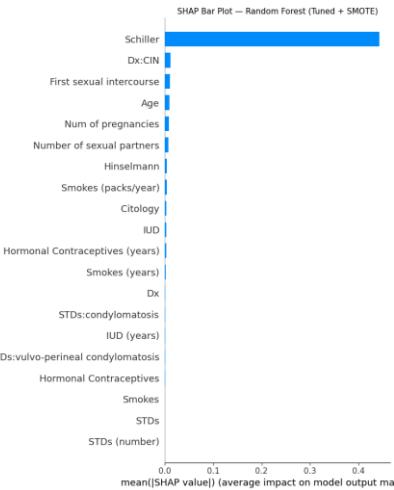
Hasil summary plot SHAP pada model Random Forest (tuned + SMOTE) ditunjukkan pada Gambar 12. Setiap titik pada grafik mewakili pengaruh suatu fitur terhadap hasil prediksi, dengan gradasi warna biru hingga merah menandakan nilai fitur dari rendah ke tinggi. Dari hasil ini, fitur Schiller's test tampak memiliki pengaruh paling besar terhadap output model, diikuti oleh Hinselmann, Citology, Jumlah kehamilan, dan STD count.

Hal ini sejalan dengan temuan Chadaga et al. [6], di mana fitur hasil tes visual seperti Schiller dan Hinselmann merupakan indikator utama keberadaan kelainan serviks.



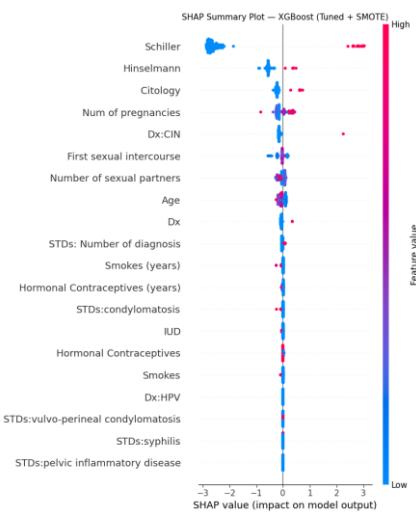
Gambar 12. SHAP Summary Plot Random Forest

Gambar 13 menunjukkan *bar plot* hasil analisis SHAP berdasarkan nilai rata-rata absolut ($mean(|SHAP|)$). Fitur hasil pemeriksaan visual memiliki kontribusi terbesar terhadap prediksi, diikuti oleh usia dan penggunaan kontrasepsi hormonal sebagai faktor sekunder [3], [13], [12]. Temuan ini sejalan dengan penelitian Toffaha et al. [13] dan Roy et al. [3], yang menegaskan bahwa model Random Forest banyak bergantung pada parameter klinis yang relevan secara medis.



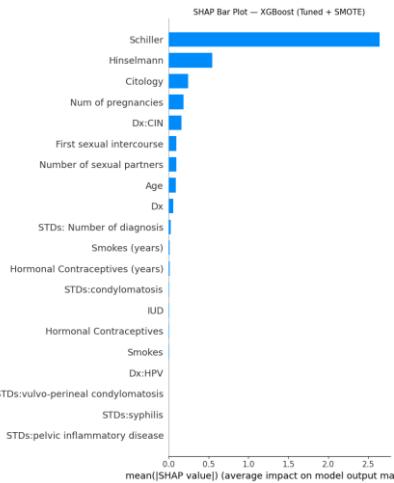
Gambar 13. SHAP Bar Plot Random Forest

Berbeda dengan Random Forest, summary plot untuk model XGBoost (tuned + SMOTE) pada Gambar 14 menunjukkan rentang nilai SHAP yang lebih lebar. Artinya, model ini lebih sensitif terhadap variasi nilai fitur dan mampu menangkap hubungan non-linear antar variabel dengan lebih baik. Fitur dominan yang muncul serupa, yakni Schiller, Hinselmann, dan Citology, yang juga ditemukan penting dalam studi Shakil et al. [22].



Gambar 14. SHAP Summary Plot XGBoost

Selain itu, pada Gambar 15, *bar plot* hasil analisis SHAP menegaskan bahwa hasil tes visual tetap menjadi fitur paling menentukan dalam keputusan model. Hasil ini menunjukkan konsistensi antar model serta kesesuaian dengan temuan penelitian sebelumnya yang mengaitkan hasil pemeriksaan Schiller dan Hinselmann dengan tingkat keparahan lesi serviks [3], [7], [13].



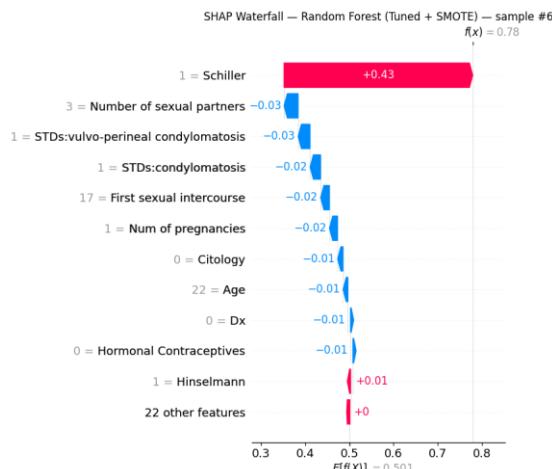
Gambar 15. SHAP Bar Plot XGBoost

2) Interpretasi Global

Untuk memahami bagaimana model membuat keputusan pada individu tertentu, digunakan SHAP waterfall plot.

Pada Gambar 9, yang menunjukkan hasil interpretasi pada model Random Forest (tuned + SMOTE), fitur Schiller = 1 menjadi kontributor terbesar dalam meningkatkan probabilitas hasil positif ($f(x) = 0.78$), diikuti oleh Hinselmann = 1 dan Number of pregnancies. Sementara itu, Age dan Cytology negatif memberikan kontribusi negatif yang menurunkan hasil prediksi.

Interpretasi ini sejalan dengan hasil penelitian Roy et al. [3], di mana variabel usia dan hasil sitologi menjadi indikator kuat dalam penentuan risiko kanker serviks.

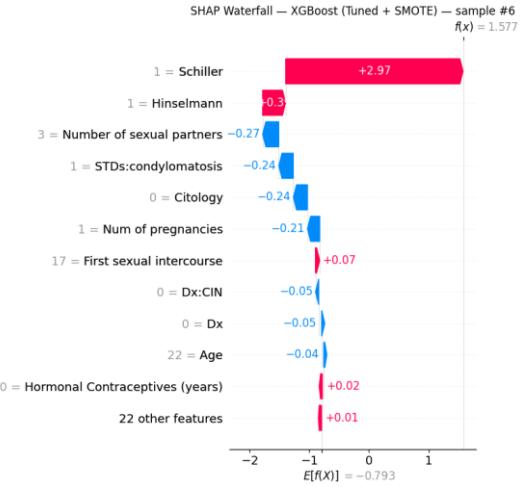


Gambar 16. SHAP Waterfall Random Forest

Sementara itu, Gambar 17 menampilkan hasil interpretasi lokal pada model XGBoost (tuned + SMOTE). Pola yang dihasilkan relatif konsisten dengan model sebelumnya, namun distribusi kontribusi tiap fitur lebih tajam. Fitur

Schiller, Hinselmann, dan Cytology kembali menjadi faktor dominan yang mendorong prediksi menuju kelas positif.

Perbedaan kecil pada urutan kontribusi menunjukkan bahwa XGBoost memiliki sensitivitas yang lebih tinggi terhadap kombinasi fitur, sebagaimana dikemukakan Chen dan Guestrin [10].



Gambar 17. SHAP Waterfall XGBoost

3) Interpretasi Klinis dan Reliabilitas

Analisis SHAP membuktikan bahwa kedua model tidak hanya akurat secara statistik, tetapi juga konsisten dengan logika medis. Fitur seperti Schiller's test, Hinselmann, dan Cytology memiliki makna klinis yang kuat dan umum digunakan dalam pemeriksaan skrining kanker serviks [3], [6]. Nilai SHAP positif menunjukkan kontribusi terhadap peningkatan risiko kanker, sedangkan nilai negatif menurunkan probabilitas diagnosis positif.

Selain itu, interpretasi dilakukan pada data uji (*test set*), bukan pada data hasil *oversampling SMOTE*, untuk menghindari bias akibat duplikasi data sintetis [7], [8]. Namun, perlu dicatat bahwa nilai SHAP bersifat asosiatif, bukan kausal, sehingga hasil interpretasi ini masih memerlukan validasi dari pakar medis sebelum digunakan dalam sistem pendukung keputusan klinis [3], [12], [13].

IV. KESIMPULAN

Penelitian ini membandingkan dua algoritma ensemble tree-based, yaitu Random Forest (RF) dan Extreme Gradient Boosting (XGBoost), dengan penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk menyeimbangkan distribusi kelas serta pendekatan SHapley Additive exPlanations (SHAP) guna meningkatkan interpretabilitas model.

Hasil pengujian menunjukkan bahwa kedua model mencapai performa klasifikasi yang sangat tinggi dengan akurasi sebesar 96,5% dan nilai Area Under Curve (AUC)

melebihi 0,89. Secara spesifik, model XGBoost (Tuned + SMOTE) memperoleh AUC sebesar 0,955, sedangkan Random Forest (Tuned + SMOTE) mencapai AUC sebesar 0,891. Nilai ini menegaskan bahwa kedua algoritma memiliki kemampuan diskriminatif yang sangat baik, dengan XGBoost sedikit lebih unggul dalam membedakan pasien berisiko tinggi dan rendah terhadap kanker serviks. Dari sisi interpretabilitas, analisis SHAP mengidentifikasi tiga fitur medis utama yang paling berpengaruh terhadap hasil prediksi, yaitu Schiller Test, Hinselmann Test, dan Cytology (Pap Smear). Ketiga fitur tersebut konsisten dengan indikator klinis yang lazim digunakan dalam diagnosis kanker serviks.

Selain itu faktor demografis seperti Age and Number of Pregnancies juga berkontribusi secara moderat terhadap peningkatan risiko, menunjukkan relevansi klinis yang sejalan dengan literatur medis. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa integrasi XGBoost, SMOTE, dan SHAP mampu menghasilkan model prediksi risiko kanker serviks yang akurat, stabil, dan dapat dijelaskan secara klinis. Pendekatan ini berpotensi diimplementasikan dalam sistem pendukung keputusan medis (Clinical Decision Support System) untuk membantu proses skrining dini kanker serviks secara lebih efisien, objektif, dan transparan.

DAFTAR PUSTAKA

- [1] World Health Organization, “Global Cancer Observatory (GLOBOCAN) 2023 Report,” 2023.
- [2] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia 2024,” 2024. [Online]. Available: <https://pusdatin.kemkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-Indonesia-2024.pdf>. [Accessed: 18-Oct-2025].
- [3] P. Roy, M. Hasan, M. R. Islam, and M. P. Uddin, “Interpretable artificial intelligence (AI) for cervical cancer risk analysis leveraging stacking ensemble and expert knowledge,” *Digit. Heal.*, vol. 11, 2025.
- [4] A. AlMohimeed, H. Saleh, S. Mostafa, R. M. A. Saad, and A. S. Talaat, “Cervical Cancer Diagnosis Using Stacked Ensemble Model and Optimized Feature Selection: An Explainable Artificial Intelligence Approach,” *Computers*, vol. 12, no. 10, 2023.
- [5] J. Fernandes, K.; Cardoso, J.; Fernandes, “Cervical Cancer (Risk Factors),” *UCI Machine Learning Repository*, 2017. [Online]. Available: <https://doi.org/10.24432/C5Z310>.
- [6] K. Chadaga, S. Prabhu, N. Sampathila, R. Chadaga, S. Swathi, and S. Sengupta, “Predicting cervical cancer biopsy results using demographic and epidemiological parameters: a custom stacked ensemble machine learning approach,” *Cogent Eng.*, vol. 9, no. 1, 2022.
- [7] M. M. Muraru, Z. Simó, and L. B. Iantovics, “Cervical Cancer Prediction Based on Imbalanced Data Using Machine Learning Algorithms with a Variety of Sampling Methods,” *Appl. Sci.*, vol. 14, no. 22, pp. 1–22, 2024.
- [8] H. Karamti *et al.*, “Improving Prediction of Cervical Cancer Using KNN Imputed SMOTE Features and Multi-Model Ensemble Learning Approach,” *Cancers (Basel)*, vol. 15, no. 17, 2023.
- [9] L. E. O. Breiman, “Random Forests,” pp. 5–32, 2001.
- [10] T. Chen and C. Guestrin, “XGBoost : A Scalable Tree Boosting System,” pp. 785–794, 2016.
- [11] B. Allen, “The Promise of Explainable AI in Digital Health for Precision Medicine: A Systematic Review,” *J. Pers. Med.*, vol. 14, no. 3, 2024.
- [12] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.
- [13] K. Toffaha, M. C. E. Simsekler, A. Slepchenko, M. A. Kortt, and L. L. Bukasa, “A Machine Learning and Bayesian Belief Network Approach to Predicting Cervical Cancer Risk: Implications for Risk Management,” *J. Multidiscip. Healthc.*, vol. 18, no. May, pp. 5199–5211, 2025.
- [14] J. Fernandes, K.; Cardoso, J.; Fernandes, “Cervical Cancer (Risk Factors).” [Online]. Available: <https://doi.org/10.24432/C5Z310>.
- [15] B. Vazquez *et al.*, “Machine and Deep Learning for the Diagnosis, Prognosis, and Treatment of Cervical Cancer: A Scoping Review,” *Diagnostics*, vol. 15, no. 12, pp. 1–42, 2025.
- [16] T. Chen and C. Guestrin, “XGBoost,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [17] H. Ahmadzadeh Sarhangi, D. Beigifard, E. Farmani, and H. Bolhasani, “Deep learning techniques for cervical cancer diagnosis based on pathology and colposcopy images,” *Informatics Med. Unlocked*, vol. 47, no. March, p. 101503, 2024.
- [18] M. S. Ali, M. M. Hossain, M. A. Kona, K. R. Nowrin, and M. K. Islam, “An ensemble classification approach for cervical cancer prediction using behavioral risk factors,” *Healthc. Anal.*, vol. 5, no. February, p. 100324, 2024.
- [19] R. Chauhan, A. Goel, B. Alankar, and H. Kaur, “Predictive modeling and web-based tool for cervical cancer risk assessment: A comparative study of machine learning models,” *MethodsX*, vol. 12, no. March, 2024.
- [20] K. M. M. Uddin, A. Al Mamun, A. Chakrabarti, R. Mostafiz, and S. K. Dey, “An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection,” *Neurosci. Informatics*, vol. 4, no. 3, p. 100169, 2024.
- [21] M. I. H. Siddiqui *et al.*, “Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI,” *Informatics Med. Unlocked*, vol. 56, no. February, p. 101657, 2025.
- [22] R. Shakil, S. Islam, and B. Akter, “A precise machine learning model: Detecting cervical cancer using feature selection and explainable AI,” *J. Pathol. Inform.*, vol. 15, no. July, p. 100398, 2024.
- [23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [24] D. Anggitasyah and M. A. P. Siregar, “Penerapan Metode Smote Extreme Gradient Boosting Untuk Klasifikasi Penyakit Kanker Serviks Di Kota Medan,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, p. 526, 2023.
- [25] Riska Chairunisa, Adiwijaya, and Widi Astuti, “Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 805–812, 2020.
- [26] L. Qian, Q. Huang, Y. Chen, and J. Chen, “A Voting-Stacking Ensemble of Inception Networks for Cervical Cytology Classification,” 2023.