

BAB 2

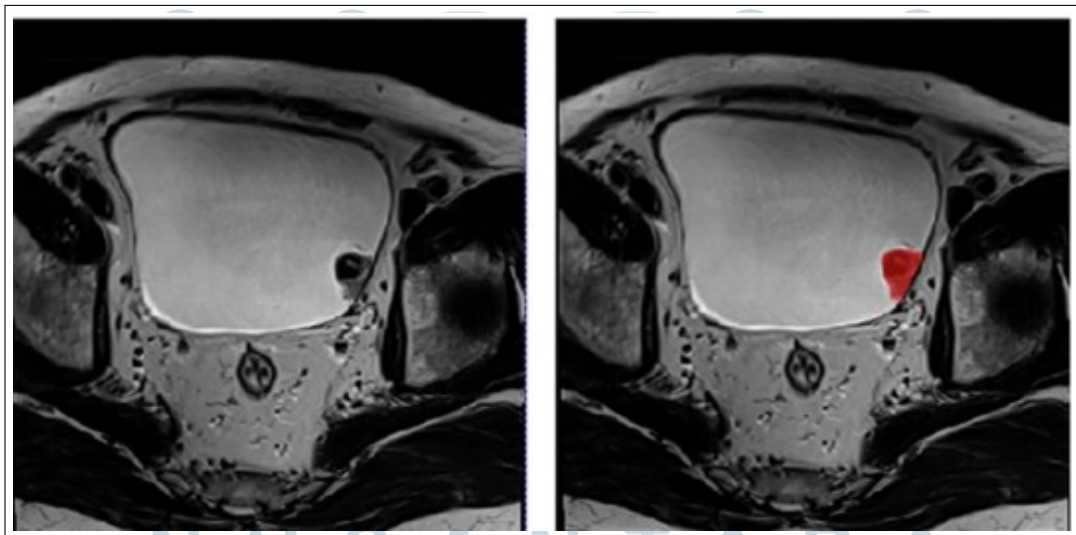
TINJAUAN PUSTAKA

Pada bab ini, akan dijabarkan teori-teori yang mendasari penelitian secara umum, berfokus pada konsep-konsep yang relevan dan mudah dipahami.

2.1 Kanker Kandung Kemih

Kanker kandung kemih adalah keganasan yang bermula dari sel-sel di lapisan dalam kandung kemih [12]. Penentuan stadium sangat penting karena menggambarkan sejauh mana kanker telah menyebar, yang menjadi dasar utama dalam perencanaan terapi [6],[7]. Secara medis, dokter membaginya menjadi dua kelompok besar berdasarkan kedalaman akarnya:

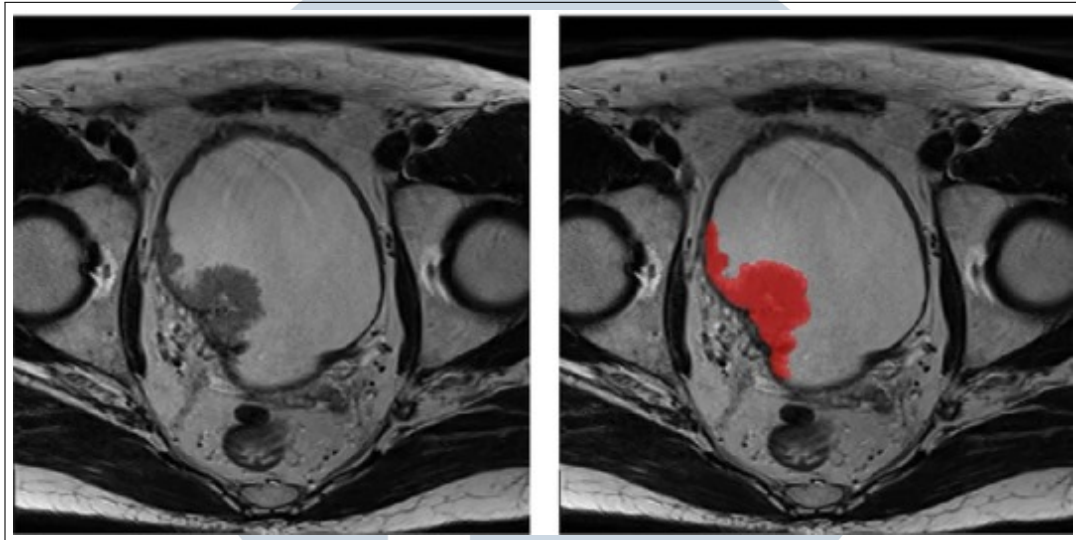
1. *NMIBC (Non-Muscle Invasive)* Ini adalah jenis yang akarnya dangkal. Sel kanker hanya tumbuh di lapisan kulit dalam atau lapisan penopang di bawahnya, tetapi belum menyentuh daging atau otot utama dinding kandung kemih. Jenis ini lebih sering ditemukan namun mudah kambuh [6]. Visualisasi contoh citra MRI pasien dengan NMIBC beserta segmentasinya dapat dilihat pada Gambar 2.1



Gambar 2.1. Contoh citra MRI dan segmentasi tumor kasus NMIBC (diadaptasi dari [1])

2. *MIBC (Muscle Invasive)*: Ini adalah jenis yang akarnya dalam. Sel kanker sudah menembus masuk lapisan otot yang tebal. Jenis ini jauh lebih

berbahaya karena pembuluh darah di otot sangat banyak, sehingga sel kanker mudah menyebar (*metastasis*) ke organ lain [7]. Visualisasi contoh citra MRI pasien dengan MIBC beserta segmentasinya dapat dilihat pada Gambar 2.2



Gambar 2.2. Contoh citra MRI dan segmentasi tumor kasus MIBC (diadaptasi dari [1])

2.2 Rekayasa Fitur (*Feature Engineering*)

Rekayasa fitur adalah proses dalam pembelajaran mesin yang bertujuan untuk mengubah data mentah menjadi fitur-fitur informatif yang dapat merepresentasikan masalah dengan lebih baik kepada model prediktif. Dalam konteks citra medis, data mentahnya adalah piksel gambar, yang perlu diubah menjadi fitur kuantitatif sebelum dapat digunakan oleh algoritma klasifikasi. Proses ini terdiri dari dua tahap utama: ekstraksi fitur dan seleksi fitur [13].

2.3 Ekstraksi Fitur (*Radiomics*)

Radiomics adalah bidang ilmu yang berfokus pada ekstraksi sejumlah besar fitur kuantitatif dari citra medis, seperti *CT scan* [14]. Fitur-fitur ini, yang seringkali tidak terlihat oleh mata manusia, dapat menangkap karakteristik tumor seperti bentuk, ukuran, dan heterogenitas internal. Proses ini dimulai dengan segmentasi *Region of Interest (ROI)*, yaitu area tumor. Setelah *ROI* ditentukan, fitur-fitur radiomics diekstraksi dan dapat dikelompokkan ke dalam beberapa kelas utama sesuai standar dari *Image Biomarker Standardisation Initiative (IBSI)* [14].

- **Fitur Bentuk (*Shape-based Features*)**

Fitur ini mengkuantifikasi properti geometris dari *ROI* dan hanya bergantung pada hasil segmentasi, bukan pada intensitas piksel di dalamnya. Fitur ini menggambarkan ukuran dan kompleksitas bentuk tumor.

- **Fitur Orde Pertama (*First-Order Statistics*)**

Fitur ini menggambarkan distribusi intensitas vokal di dalam *ROI* berdasarkan histogram, tanpa mempertimbangkan hubungan spasial antar vokal.

- ***Gray Level Co-occurrence Matrix (GLCM)***

GLCM adalah matriks yang mengukur hubungan spasial antara pasangan vokal dengan intensitas tertentu pada jarak dan arah yang telah ditentukan. Matriks ini menjadi dasar untuk fitur tekstur yang menggambarkan homogenitas dan kontras lokal. Beberapa fitur yang diekstraksi dari *GLCM* antara lain *Contrast*, *Correlation*, *Energy*, dan *Homogeneity* [14].

- ***Gray Level Run Length Matrix (GLRLM)***

GLRLM mengkuantifikasi deretan vokal berurutan yang memiliki tingkat keabuan yang sama dalam arah tertentu. Fitur ini berguna untuk mengidentifikasi tekstur kasar atau halus. Matriks ini menghitung jumlah *run* (deretan) dengan panjang tertentu untuk setiap tingkat keabuan [14].

- ***Gray Level Size Zone Matrix (GLSZM)***

GLSZM mengkuantifikasi zona-zona vokal yang terhubung dan memiliki tingkat keabuan yang sama. Sebuah zona adalah area 3D di mana semua vokal memiliki intensitas yang sama dan saling terhubung. Fitur ini menggambarkan homogenitas regional dan tidak bergantung pada arah [14].

- ***Gray Level Dependence Matrix (GLDM)***

GLDM mengkuantifikasi dependensi tingkat keabuan di dalam sebuah lingkungan. Sebuah vokal dianggap dependen pada vokal pusat jika perbedaan intensitasnya berada di bawah ambang batas tertentu. Matriks ini menghitung jumlah dependensi untuk setiap tingkat keabuan [14].

- ***Neighbouring Gray Tone Difference Matrix (NGTDM)***

NGTDM mengukur perbedaan antara intensitas sebuah vokal dengan rata-rata intensitas tetangganya. Matriks ini menghasilkan fitur seperti

Coarseness, *Contrast*, dan *Busyness*, yang menggambarkan seberapa cepat intensitas berubah dalam sebuah lingkungan lokal [14].

2.4 Seleksi Fitur

Setelah proses ekstraksi, seringkali dihasilkan ratusan atau ribuan fitur. Banyak dari fitur ini mungkin tidak relevan atau redundan, yang dapat menurunkan performa model. Seleksi fitur adalah proses untuk memilih sub-set fitur yang paling informatif dan relevan untuk tugas prediksi. Dalam penelitian ini, digunakan algoritma *Boruta*.

Boruta adalah algoritma seleksi fitur yang bekerja berdasarkan prinsip semua relevan (*all-relevant*), yang bertujuan untuk menemukan semua fitur yang memiliki hubungan prediktif dengan variabel target [15]. Cara kerjanya secara konseptual adalah sebagai berikut:

1. Pembuatan Fitur Bayangan (*Shadow Features*): *Boruta* membuat salinan dari seluruh dataset fitur, lalu mengacak nilai di setiap kolom salinan tersebut. Fitur-fitur acak ini disebut fitur bayangan (*shadow features*) dan berfungsi sebagai referensi acak.
2. Pelatihan Model: *Dataset* asli digabungkan dengan fitur bayangan, dan sebuah model *classifier* (*Random Forest*) dilatih pada *dataset* gabungan ini untuk menghitung *feature importance* dari semua fitur (asli dan bayangan).
3. Perbandingan dan Keputusan: Kepentingan setiap fitur asli dibandingkan dengan kepentingan fitur bayangan terbaik. Sebuah fitur asli dianggap penting jika kepentingannya secara konsisten lebih tinggi daripada fitur bayangan terbaik setelah melalui beberapa iterasi. Fitur yang kepentingannya lebih rendah akan ditolak.

2.5 Voxel dan Peran Isotropic Resampling pada MRI untuk Radiomik

Dalam pencitraan medis dan pemrosesan citra tiga dimensi, *voxel* (*volume element*) merupakan unit dasar yang digunakan untuk merepresentasikan informasi spasial dalam ruang tiga dimensi. *Voxel* dapat dipandang sebagai ekstensi piksel pada citra dua dimensi, di mana setiap *voxel* merepresentasikan nilai intensitas atau atribut tertentu pada suatu posisi di dalam volume citra. Representasi berbasis voxel memungkinkan struktur anatomi dan objek medis dimodelkan secara diskret dalam

ruang 3D, sehingga mendukung analisis visual maupun kuantitatif pada berbagai modalitas pencitraan medis seperti *Computed Tomography* (CT) dan *Magnetic Resonance Imaging* (MRI) [16], [17]. Ukuran voxel secara langsung menentukan resolusi spasial citra dan tingkat detail informasi yang dapat direpresentasikan dalam analisis medis.

Pencitraan medis berbasis *MRI*, adalah citra tiga dimensi tersusun atas elemen volumetrik yang disebut *voxel*, yang merepresentasikan resolusi spasial pada arah sumbu x , y , dan z . Perbedaan parameter akuisisi citra, jenis pemindai, serta protokol pencitraan antar pusat layanan kesehatan dapat menyebabkan variasi ukuran *voxel*, khususnya pada dataset *MRI multi-center* [18]. Variasi resolusi *voxel* ini berpotensi menimbulkan heterogenitas spasial yang dapat memengaruhi konsistensi dan stabilitas fitur radiomik yang diekstraksi. Penelitian terdahulu menunjukkan bahwa perubahan parameter akuisisi yang secara langsung memengaruhi ukuran *voxel*, seperti *pixel spacing* dan *slice thickness*, dapat menyebabkan variasi signifikan pada nilai serta *repeatability* fitur radiomik, terutama pada fitur tekstur tiga dimensi [19].

Isotropic voxel resampling merupakan salah satu teknik praproses citra yang bertujuan untuk menyeragamkan ukuran *voxel* pada seluruh citra dengan menjadikan resolusi spasial sama pada setiap dimensi. Dengan menerapkan *resampling* isotropik, citra *MRI* dengan resolusi awal yang berbeda dapat distandarisasi ke ukuran *voxel* tertentu, sehingga mengurangi bias akibat perbedaan resolusi dan meningkatkan keterbandingan antar citra. Pendekatan ini banyak digunakan dalam studi radiomik untuk meningkatkan reproduisibilitas dan keandalan fitur yang dihasilkan [11].

Dalam konteks radiomik, *resampling voxel* umumnya dilakukan sebelum proses ekstraksi fitur, karena banyak fitur tekstur dan bentuk sangat sensitif terhadap perubahan resolusi spasial. Standarisasi ukuran *voxel* diharapkan dapat meningkatkan stabilitas fitur serta kinerja model pembelajaran mesin yang dibangun berdasarkan fitur-fitur tersebut. Oleh karena itu, *isotropic voxel resampling* menjadi tahapan penting dalam pipeline radiomik, khususnya pada penelitian yang melibatkan data *MRI* dari berbagai pusat dengan variasi protokol pencitraan.

2.6 Algoritma Pembelajaran Mesin (*Machine Learning*)

Penelitian ini menerapkan tiga algoritma utama untuk klasifikasi status invasi otot, yang masing-masing mewakili paradigma pembelajaran yang berbeda: geometris (*Support Vector Machine*), *ensemble* (*eXtreme Gradient Boosting*), dan koneksionis (*MLP (Multilayer Perceptron)/Neural Network*). Pemilihan ketiga algoritma ini didasarkan pada efektivitasnya yang terbukti dalam studi radiomik kanker kandung kemih terkini [20], [21].

2.6.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang bekerja dengan prinsip pencarian *hyperplane* (bidang pemisah) optimal yang memisahkan dua kelas data dengan margin terbesar. SVM dikenal memiliki kemampuan generalisasi yang baik pada dataset dengan dimensi tinggi namun jumlah sampel terbatas, seperti yang sering ditemukan dalam radiomik [22]. Cara kerjanya secara konseptual adalah sebagai berikut:

1. Konsep *Hyperplane* dan *Margin*

Diberikan sekumpulan data latih (x_i, y_i) di mana x_i adalah vektor fitur radiomics dan $y_i \in \{-1, +1\}$ adalah label kelas (*NMIBC/MIBC*). SVM mencari vektor bobot w dan bias b yang memenuhi persamaan *hyperplane*:

$$w \cdot x + b = 0 \quad (2.1)$$

Tujuan utamanya adalah memaksimalkan jarak (*margin*) antara *hyperplane* dan titik data terdekat (*support vectors*), yang secara matematis dirumuskan sebagai meminimalkan $\|w\|^2$ dengan batasan:

$$y_i(w \cdot x_i + b) \geq 1 \quad (2.2)$$

2. *Kernel Trick*

Karena data *radiomics* seringkali tidak dapat dipisahkan secara linier (*non-linearly separable*), penelitian ini menggunakan fungsi *kernel Radial Basis Function (RBF)* untuk memetakan data ke dimensi yang lebih tinggi agar

dapat dipisahkan. Fungsi *kernel RBF* didefinisikan sebagai:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.3)$$

Parameter γ mengontrol seberapa jauh pengaruh satu sampel pelatihan, sedangkan parameter C (regularisasi) mengontrol *trade-off* antara kesalahan klasifikasi dan lebar *margin*. Optimasi parameter ini krusial untuk mencegah *overfitting* [22].

2.6.2 eXtreme Gradient Boosting (XGBoost)

XGBoost adalah pengembangan dari algoritma *Gradient Boosting Decision Tree (GBDT)* yang dirancang untuk kecepatan dan performa tinggi pada data tabular. Dalam studi radiomik kanker kandung kemih, *XGBoost* sering menunjukkan performa superior karena kemampuannya menangani interaksi fitur non-linier dan data yang tidak seimbang [23]. Cara kerjanya secara konseptual adalah sebagai berikut:

1. Prinsip Boosting

Berbeda dengan model tunggal, *XGBoost* membangun model secara bertahap (*additive*). Model memprediksi hasil dengan menjumlahkan skor dari sekumpulan pohon keputusan (f_k):

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (2.4)$$

Di mana setiap pohon baru ditambahkan untuk memperbaiki kesalahan (*residual*) dari pohon sebelumnya.

2. Fungsi Objektif dan Regularisasi

Keunggulan utama *XGBoost* dibanding *GBDT* biasa adalah adanya komponen regularisasi (Ω) dalam fungsi objektifnya untuk mencegah *overfitting*, yang sangat krusial pada *dataset radiomics* dengan jumlah fitur banyak. Fungsi objektif yang diminimalkan adalah:

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2.5)$$

Di mana l adalah fungsi *loss* (misalnya log-loss untuk klasifikasi) dan $\Omega(f)$ adalah penalti kompleksitas pohon yang mencakup jumlah daun dan bobot daun:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda ||w||^2 \quad (2.6)$$

Fitur ini membuat *XGBoost* sangat efektif dalam menangkap pola non-linier sekaligus menjaga generalisasi model [23].

2.6.3 Multilayer Perceptron - MLP (Neural Network)

Multilayer Perceptron (MLP) adalah jenis jaringan saraf tiruan yang terdiri dari lapisan input, satu atau lebih lapisan tersembunyi (*hidden layers*), dan lapisan output. *MLP* telah digunakan dalam model *Deep Radiomics* untuk menangkap representasi fitur abstrak yang mendalam dari citra *CT* dan *MRI* [20].

Setiap *neuron* dalam lapisan tersembunyi melakukan transformasi linier diikuti oleh fungsi aktivasi non-linier. Jika x adalah input, W adalah bobot, dan b adalah bias, output neuron (h) dihitung sebagai:

$$h(x) = f(W \cdot x + b) \quad (2.7)$$

Penelitian ini menggunakan fungsi aktivasi *ReLU* (*Rectified Linear Unit*) ($f(x) = \max(0, x)$) pada lapisan tersembunyi untuk mempercepat konvergensi, dan fungsi *Sigmoid* pada lapisan output untuk menghasilkan probabilitas prediksi biner (0 hingga 1)[20].

2.6.4 Hyperparameter Tuning dengan GridSearchCV

Hyperparameter tuning merupakan tahapan yang sangat penting dalam proses pengembangan model pembelajaran mesin karena berpengaruh secara signifikan terhadap kinerja serta kemampuan generalisasi model. Berbagai metode penalaan telah dikembangkan dan digunakan, antara lain *grid search*, *random search*, optimisasi *Bayesian*, dan *meta-learning* [24]. Pada penelitian ini, proses penalaan hyperparameter dilakukan menggunakan metode *GridSearchCV*, yang mengevaluasi seluruh kombinasi hyperparameter dalam ruang pencarian yang telah ditentukan untuk memperoleh konfigurasi model yang optimal [25].

GridSearchCV merupakan metode pencarian hyperparameter berbasis *exhaustive search* yang mengevaluasi seluruh kombinasi parameter dalam

ruang pencarian yang telah ditentukan. Proses evaluasi dilakukan dengan mengombinasikan *k-fold cross-validation* untuk mengurangi bias akibat pembagian data latih dan validasi yang bersifat acak. Pada penelitian ini digunakan *5-fold cross-validation*, di mana data dibagi menjadi lima subset dan setiap subset digunakan secara bergantian sebagai data validasi.

Secara matematis, *GridSearchCV* bertujuan untuk menentukan kombinasi hyperparameter optimal θ^* yang meminimalkan nilai fungsi loss rata-rata pada seluruh fold validasi, yang dirumuskan sebagai:

$$\theta^* = \arg \min_{\theta \in \Theta} \frac{1}{K} \sum_{k=1}^K L_k(\theta) \quad (2.8)$$

di mana Θ menyatakan ruang pencarian hyperparameter, K adalah jumlah fold pada *cross-validation*, dan $L_k(\theta)$ merupakan nilai fungsi loss pada fold ke- k .

Dalam penelitian ini, *GridSearchCV* diterapkan untuk mengoptimalkan hyperparameter pada setiap algoritma klasifikasi. Pada model *Support Vector Machine (SVM)*, parameter yang dituning meliputi parameter regularisasi C , parameter *kernel* γ , serta jenis *kernel* yang digunakan. Untuk model *eXtreme Gradient Boosting (XGBoost)*, proses tuning mencakup jumlah *estimator*, kedalaman maksimum pohon, *learning rate*, rasio *subsampling*, dan parameter γ untuk mengontrol kompleksitas pohon. Sementara itu, pada model *Multilayer Perceptron (MLP)*, parameter yang dioptimalkan meliputi jumlah dan konfigurasi *neuron* pada *hidden layer*, fungsi aktivasi, parameter regularisasi α , serta skema *learning rate*. Pendekatan ini memastikan bahwa setiap algoritma dievaluasi pada kondisi parameter terbaiknya masing-masing, sehingga perbandingan performa model menjadi lebih objektif dan reliabel.

2.7 Metrik Evaluasi

Untuk mengevaluasi performa model secara terpisah, digunakan beberapa metrik yang dapat menggambarkan kinerja model dari berbagai aspek, terutama dalam menganalisis pola kesalahan yang terjadi.

Confusion Matrix adalah sebuah tabel yang digunakan untuk memvisualisasikan performa dari sebuah algoritma klasifikasi. Dalam konteks klasifikasi multi-kelas dengan N kelas, matriks ini akan berukuran $N \times N$. Setiap baris merepresentasikan kelas aktual, sementara setiap kolom merepresentasikan kelas yang diprediksi oleh model [26]. Struktur matriks ini memungkinkan analisis

mendalam terhadap pola kesalahan prediksi dan menjadi dasar untuk menghitung metrik-metrik evaluasi lainnya.

Untuk mengukur seberapa baik model klasifikasi membedakan antara kelas *NMIBC* dan *MIBC*, digunakan *Confusion Matrix*. Matriks ini merepresentasikan perbandingan antara prediksi model dengan label kebenaran dasar (*ground truth*). Berdasarkan matriks ini, empat parameter dasar dapat dihitung:

1. *True Positive* (TP): Jumlah kasus positif yang diprediksi dengan benar (misal: *MIBC* terdeteksi sebagai *MIBC*).
2. *True Negative* (TN): Jumlah kasus negatif yang diprediksi dengan benar (misal: *NMIBC* terdeteksi sebagai *NMIBC*).
3. *False Positive* (FP): Jumlah kasus negatif yang salah diprediksi sebagai positif (Error Tipe I).
4. *False Negative* (FN): Jumlah kasus positif yang salah diprediksi sebagai negatif (Error Tipe II).

Berdasarkan parameter tersebut, kinerja model diukur menggunakan metrik-metrik berikut:

Accuracy mengukur proporsi total prediksi yang benar terhadap seluruh jumlah sampel data. Metrik ini memberikan gambaran umum mengenai seberapa sering model membuat keputusan yang tepat secara global. Rumus akurasi didefinisikan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

Meskipun akurasi adalah metrik yang paling intuitif, penggunaannya pada data medis memerlukan kehati-hatian. Dalam dataset yang tidak seimbang (*imbalanced dataset*), di mana jumlah pasien satu kelas jauh lebih banyak dari kelas lainnya, akurasi yang tinggi bisa menipu (*misleading*). Model dapat mencapai akurasi tinggi hanya dengan memprediksi kelas mayoritas saja, namun gagal mendeteksi kasus minoritas yang mungkin justru lebih klinis dan kritis [26].

Precision atau disebut juga *Positive Predictive Value*, mengukur seberapa andal model saat memprediksi kelas positif. Metrik ini menghitung dari semua

pasien yang diprediksi mengidap kanker ganas (*MIBC*) berapa persen yang benar-benar mengidapnya.

$$Precision = \frac{TP}{TP + FP} \quad (2.10)$$

Nilai presisi yang tinggi mengindikasikan tingkat *False Positive* (FP) yang rendah. Dalam konteks medis, presisi sangat penting untuk meminimalkan alarm palsu, di mana pasien yang sebenarnya kondisinya ringan tidak perlu menjalani prosedur lanjutan yang invasif dan memicu kecemasan yang tidak perlu [15].

Recall dalam dunia medis lebih dikenal sebagai Sensitivitas (*Sensitivity*), mengukur kemampuan model untuk menemukan semua kasus positif yang ada. Metrik ini menghitung dari seluruh pasien yang benar-benar mengidap kanker ganas, berapa persen yang berhasil dideteksi oleh sistem.

$$Recall = \frac{TP}{TP + FN} \quad (2.11)$$

Dalam diagnosis kanker, *Recall* sering dianggap sebagai metrik yang paling krusial. Nilai *Recall* yang rendah berarti banyak terjadi *False Negative* (FN), yaitu pasien yang sakit tidak terdeteksi. Kesalahan ini dapat berakibat fatal karena pasien kehilangan kesempatan untuk mendapatkan penanganan dini yang menyelamatkan nyawa [15].

F1-Score adalah rata-rata harmonik dari Presisi dan *Recall*. Berbeda dengan rata-rata aritmatika biasa, rata-rata harmonik memberikan bobot lebih pada nilai yang rendah, sehingga *F1-Score* hanya akan bernilai tinggi jika kedua komponen (Presisi dan *Recall*) seimbang dan tinggi.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.12)$$

F1-Score menjadi metrik pilihan utama dalam penelitian klasifikasi medis, terutama ketika *dataset* memiliki distribusi kelas yang tidak seimbang. Metrik ini memberikan penilaian yang lebih objektif dibandingkan akurasi karena menuntut model untuk memiliki kinerja yang baik dalam meminimalisir kesalahan positif palsu maupun negatif palsu secara bersamaan [15].