BAB 2
LANDASAN TEORI

2.1 Kanker Payudara

Kanker payudara merupakan jenis kanker yang paling sering didiagnosis pada perempuan di seluruh dunia. Kanker ini juga menjadi penyebab utama kematian pada wanita akibat kanker. Berdasarkan laporan Global Cancer Observatory (GLOBOCAN), kanker payudara menempati peringkat pertama dalam jumlah kasus baru secara global, melampaui jenis kanker lainnya pada perempuan, serta berkontribusi signifikan terhadap angka mortalitas akibat kanker [1]. Tingginya prevalensi ini menjadikan kanker payudara sebagai salah satu isu kesehatan global yang memerlukan perhatian serius, khususnya dalam upaya pencegahan dan deteksi dini.

Secara klinis dan patologis, kelainan pada jaringan payudara dapat diklasifikasikan menjadi dua kategori utama, yaitu lesi jinak (*benign*) dan lesi ganas (*malignant*). Lesi jinak merupakan pertumbuhan sel abnormal yang tidak bersifat kanker, tidak invasif, dan tidak menyebar ke jaringan lain. Sebaliknya, lesi ganas ditandai oleh pertumbuhan sel yang tidak terkendali, bersifat invasif, serta memiliki kemampuan untuk bermetastasis ke organ lain melalui sistem limfatik maupun aliran darah. Perbedaan karakteristik antara kedua jenis lesi ini seringkali menjadi tantangan dalam proses diagnosis, bahkan dapat menimbulkan variasi interpretasi di antara tenaga medis [20].

Deteksi dini kanker payudara memiliki peran yang sangat penting dalam meningkatkan peluang kesembuhan pasien serta menurunkan angka kematian. Diagnosis pada stadium awal memungkinkan dilakukannya intervensi medis yang lebih efektif dan minim risiko, seperti pembedahan konservatif, terapi radiasi, maupun kemoterapi. Studi menunjukkan bahwa deteksi dini secara signifikan berkontribusi terhadap peningkatan *survival rate* pasien kanker payudara, sehingga berbagai program skrining terus dikembangkan di berbagai negara [21].

Dalam praktik klinis, proses diagnosis kanker payudara didukung oleh berbagai teknik pencitraan medis. Modalitas yang umum digunakan meliputi mammografi, *Magnetic Resonance Imaging* (MRI), dan *ultrasound*.

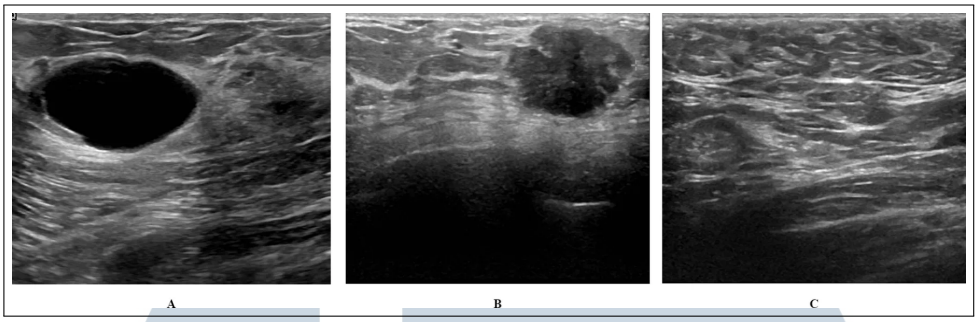
2.2 Citra Ultrasound Payudara

Ultrasound atau ultrasonografi merupakan salah satu modalitas pencitraan medis yang umum digunakan dalam diagnosis kanker payudara. Metode ini memanfaatkan gelombang suara frekuensi tinggi untuk menghasilkan gambaran struktur internal jaringan tubuh secara *real-time*. Keunggulan utama *ultrasound* terletak pada sifatnya yang non-invasif, tidak menggunakan radiasi ionisasi, serta relatif lebih aman dan ekonomis dibandingkan dengan modalitas pencitraan lainnya seperti mammografi dan *Magnetic Resonance Imaging* (MRI). Hal ini menjadikan *ultrasound* sebagai pilihan yang populer, terutama untuk pemeriksaan lanjutan maupun skrining pada kelompok pasien tertentu [22].

Dalam konteks klinis, *ultrasound* memiliki peran penting dalam mendeteksi dan mengkarakterisasi lesi pada jaringan payudara, khususnya pada pasien dengan densitas jaringan payudara yang tinggi. Jaringan payudara padat umumnya ditemukan pada perempuan usia muda dan dapat mengurangi sensitivitas mammografi dalam mendeteksi kelainan. Dalam kondisi ini, *ultrasound* menjadi modalitas pelengkap yang efektif karena mampu membedakan antara lesi kistik (berisi cairan) dan lesi padat, serta memberikan informasi tambahan mengenai bentuk, ukuran, dan batas lesi [23]. Oleh karena itu, *ultrasound* sering digunakan sebagai alat bantu diagnosis untuk meningkatkan akurasi deteksi kanker payudara. Contoh citra *ultrasound* payudara yang menunjukkan karakteristik visual tersebut dapat dilihat pada Gambar 2.1.

Meskipun memiliki berbagai keunggulan, citra *ultrasound* juga memiliki karakteristik visual yang kompleks dan menantang untuk dianalisis. Salah satu permasalahan utama adalah keberadaan *speckle noise*, yaitu *noise granular* yang muncul akibat interferensi gelombang suara selama proses akuisisi citra. *Speckle noise* ini dapat menurunkan kualitas citra dan mengaburkan detail penting pada jaringan, sehingga menyulitkan proses identifikasi lesi. Selain itu, citra *ultrasound* umumnya memiliki kontras yang rendah serta batas lesi yang tidak selalu jelas, terutama pada kasus lesi ganas yang memiliki bentuk tidak beraturan dan menyatu dengan jaringan sekitarnya [24].

UNIVERSITAS MULTIMEDIA NUSANTARA



Gambar 2.1. Contoh citra *ultrasound* payudara: (a) lesi jinak (*benign*), (b) lesi ganas (*malignant*), (c) jaringan normal (*normal*)

Kondisi tersebut membuat interpretasi citra *ultrasound* sangat dipengaruhi oleh pengalaman dan keahlian radiolog. Perbedaan interpretasi antar-observer dapat terjadi dan berpotensi memengaruhi konsistensi serta akurasi diagnosis. Oleh sebab itu, dibutuhkan pendekatan berbasis teknologi yang dapat membantu analisis citra secara lebih objektif dan konsisten. Dalam beberapa tahun terakhir, metode berbasis kecerdasan buatan (*Artificial Intelligence*), terutama *deep learning*, telah banyak dikembangkan untuk menangani permasalahan tersebut.

2.3 *Deep Learning* dalam Analisis Citra Medis

Deep learning merupakan salah satu cabang kecerdasan buatan (*Artificial Intelligence*) yang menggunakan jaringan saraf tiruan berlapis (*deep neural networks*) untuk mempelajari representasi data secara otomatis dan hierarkis. Pendekatan ini memungkinkan model mengekstraksi fitur langsung dari data mentah tanpa perlu perancangan fitur manual, yang menjadi salah satu keterbatasan pada metode *machine learning* tradisional [25].

Dalam analisis citra medis, *deep learning* telah menunjukkan performa yang sangat baik pada berbagai tugas, seperti klasifikasi, deteksi, dan segmentasi. Kemampuannya dalam mempelajari pola kompleks serta hubungan *non-linear* dari data citra membuat metode ini efektif untuk mengidentifikasi karakteristik penyakit yang sulit diamati secara visual oleh manusia. Hal ini relevan pada citra medis seperti *ultrasound*, yang memiliki kompleksitas tinggi serta sering mengandung *noise* dan variasi kontras [26].

Salah satu keunggulan utama *deep learning* adalah kemampuannya dalam melakukan *end-to-end learning*, di mana seluruh proses mulai dari ekstraksi fitur hingga klasifikasi dilakukan dalam satu model terpadu. Pendekatan ini tidak hanya meningkatkan efisiensi, tetapi juga memungkinkan model untuk

menghasilkan representasi fitur yang lebih optimal dibandingkan dengan metode berbasis fitur manual. Selain itu, dengan dukungan dataset yang semakin besar dan peningkatan kemampuan komputasi, performa *deep learning* dalam bidang medis terus mengalami peningkatan signifikan [27].

Meskipun demikian, penerapan *deep learning* dalam citra medis juga menghadapi beberapa tantangan, seperti kebutuhan data berlabel dalam jumlah besar, potensi *overfitting*, serta kurangnya transparansi dalam pengambilan keputusan model. Oleh karena itu, penelitian di bidang ini tidak hanya berfokus pada peningkatan akurasi, tetapi juga pada efisiensi model dan interpretabilitas hasil, agar dapat lebih mudah diintegrasikan dalam praktik klinis.

2.4 *Convolutional Neural Network*

Convolutional Neural Network (CNN) merupakan salah satu arsitektur *deep learning* yang dirancang untuk memproses data berbentuk *grid*, seperti citra. CNN banyak digunakan dalam analisis citra medis karena mampu mengekstraksi fitur spasial secara otomatis dan efektif. Berbeda dengan metode tradisional yang bergantung pada fitur manual, CNN dapat mempelajari representasi fitur langsung dari data selama proses pelatihan [28].

Arsitektur CNN memiliki beberapa komponen utama, yaitu *convolutional layer*, *activation function*, *pooling layer*, dan *fully connected layer*. *Convolutional layer* digunakan untuk mengekstraksi fitur citra melalui kernel atau filter yang bergerak pada seluruh area citra. Operasi konvolusi ini memungkinkan model mengenali pola lokal, seperti tepi, tekstur, dan bentuk objek, sebagaimana ditunjukkan pada Rumus 2.1.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m,\ j+n) \cdot K(m,n) \tag{2.1}$$

Setelah proses konvolusi, fungsi aktivasi diterapkan untuk memperkenalkan non-linearitas ke dalam model. Salah satu fungsi aktivasi yang paling umum digunakan adalah *Rectified Linear Unit* (ReLU), sebagaimana ditunjukkan pada Rumus 2.2.

$$\mathrm{ReLU}(x) = \max(0,x) \tag{2.2}$$

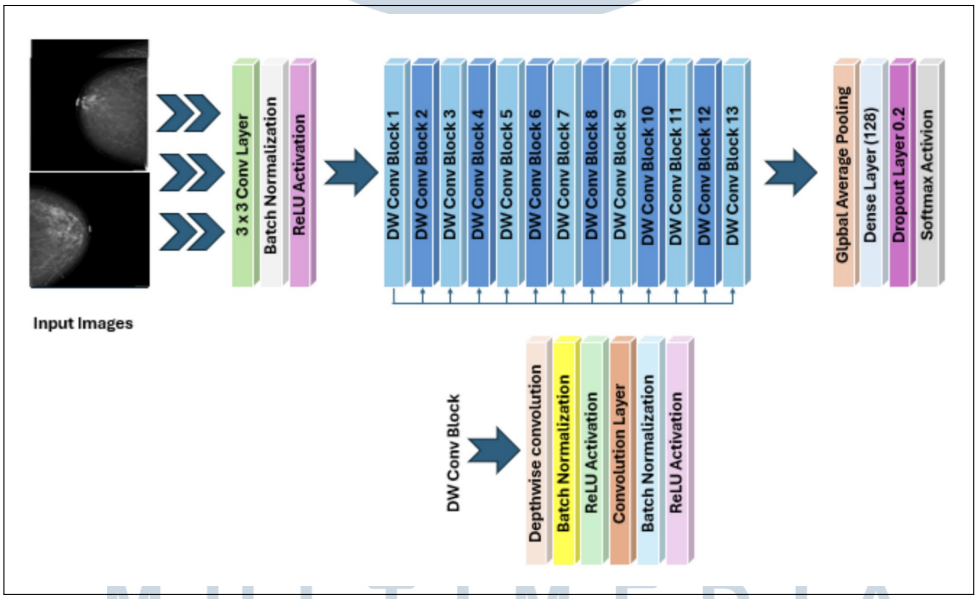
Klasifikasi Lesi Payudara..., Russel Shivah Budiarto, Universitas Multimedia Nusantara

Selanjutnya, *pooling layer* digunakan untuk mengurangi dimensi spasial dari *feature map*, sehingga dapat menurunkan kompleksitas komputasi dan mengurangi risiko *overfitting*. Teknik yang umum digunakan adalah *max pooling*, yang mengambil nilai maksimum dari suatu area tertentu pada *feature map*.

Pada tahap akhir, fitur yang telah diekstraksi diproses oleh *fully connected layer* untuk menghasilkan output klasifikasi. Salah satu keunggulan utama CNN adalah kemampuannya dalam memanfaatkan *parameter sharing* dan *local connectivity*, yang membuat model lebih efisien dibandingkan jaringan saraf tradisional [29]. Namun, CNN memiliki keterbatasan dalam menangkap hubungan global antar bagian citra, yang kemudian mendorong pengembangan arsitektur yang lebih canggih seperti *Vision Transformer*.

2.5 MobileNet

MobileNet adalah salah satu arsitektur CNN yang memiliki rancangan model yang ringan dan efisien, sehingga dapat digunakan pada perangkat dengan keterbatasan sumber daya komputasi [30].



Gambar 2.2. Struktur arsitektur *MobileNet*

Ciri khas utama dari MobileNet adalah pemanfaatan *depthwise separable convolution*. Teknik ini bekerja dengan memisahkan operasi konvolusi standar menjadi dua tahap: *depthwise convolution* dan *pointwise convolution*. Ilustrasi arsitektur dapat dilihat pada Gambar 2.2. Perbandingan kompleksitas komputasi

antara konvolusi standar dan *depthwise separable convolution* ditunjukkan pada Rumus 2.3 hingga Rumus 2.5.

$$C_{Std} = D_K^2 \cdot M \cdot N \cdot D_F^2 \tag{2.3}$$

$$C_{DS} = D_K^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2 \tag{2.4}$$

$$Computational\ Cost\ Ratio = \frac{1}{N} + \frac{1}{D_K^2} \tag{2.5}$$

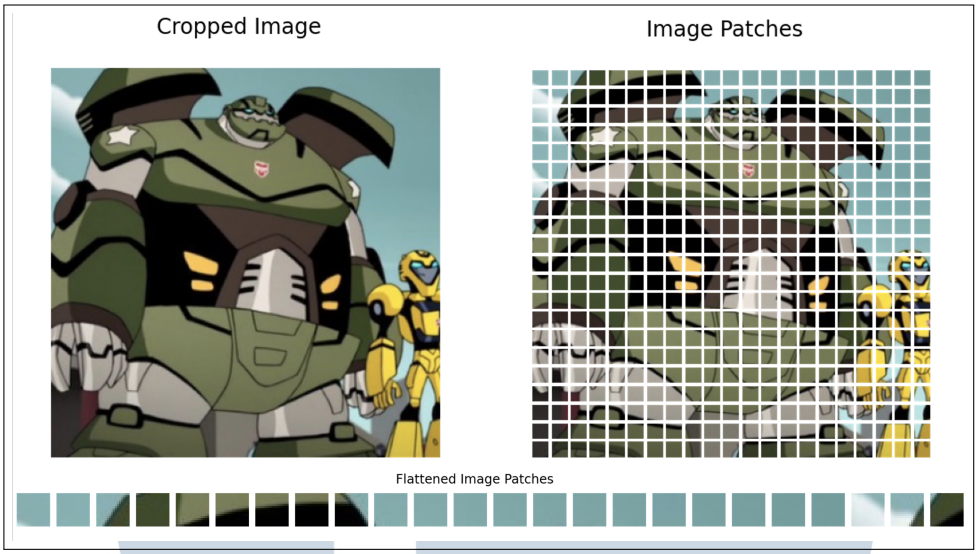
Seiring perkembangannya, MobileNet mengalami beberapa peningkatan. MobileNetV2 memperkenalkan konsep *inverted residual* dan *linear bottleneck* untuk meningkatkan efisiensi representasi fitur [31]. Pada penelitian ini, MobileNetV2 digunakan sebagai salah satu *backbone* dalam arsitektur *hybrid* karena kemampuannya mengekstraksi fitur lokal secara efisien.

2.6 *Vision Transformer* (ViT)

Vision Transformer (ViT) adalah arsitektur *deep learning* yang mengadaptasi konsep *Transformer* dari pemrosesan bahasa alami (*natural language processing*) ke domain pengolahan citra (*image processing*). Berbeda dengan CNN yang mengekstraksi fitur secara lokal, *Vision Transformer* menggunakan mekanisme *self-attention* untuk mempelajari hubungan global antar bagian citra [32].

Dalam *Vision Transformer*, citra input terlebih dahulu dibagi menjadi sejumlah *patch* berukuran tetap. Setiap *patch* kemudian di-*flatten* dan diproyeksikan menjadi vektor *embedding*, sebagaimana ditunjukkan pada Gambar 2.3.

UNIVERSITAS MULTIMEDIA NUSANTARA



Gambar 2.3. Contoh pembagian citra input menjadi *patches*

Selanjutnya, *embedding* dari setiap *patch* diproses oleh *encoder Transformer* yang terdiri dari beberapa lapisan *multi-head self-attention* dan *feed-forward network*. Mekanisme *self-attention* ditunjukkan pada Rumus 2.6.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.6}$$

Salah satu keunggulan utama *Vision Transformer* adalah kemampuannya dalam menangkap dependensi jarak jauh (*long-range dependencies*) dalam citra, yang sulit dicapai oleh CNN. Namun, ViT memiliki keterbatasan dalam menangkap fitur lokal secara detail, sehingga sering dikombinasikan dengan CNN untuk menghasilkan model yang lebih seimbang [33].

2.7 Arsitektur *Hybrid* CNN–*Vision Transformer*

Arsitektur *hybrid* MobileNet–*Vision Transformer* merupakan pendekatan yang menggabungkan keunggulan CNN dan *Transformer* dalam satu kerangka kerja terpadu untuk mengatasi keterbatasan masing-masing metode.

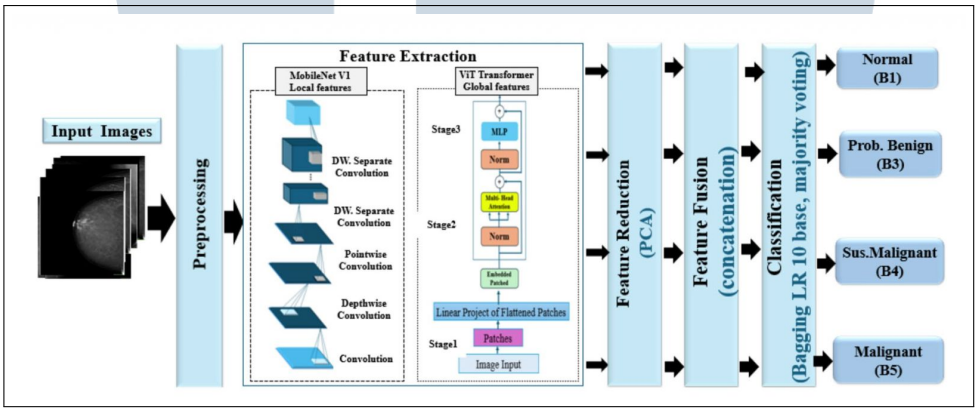
Dalam penelitian acuan yang menjadi dasar pengembangan penelitian ini, arsitektur *hybrid* diimplementasikan dalam bentuk *dual-stream framework* di mana MobileNet dan *Vision Transformer* bekerja secara paralel sebagai dua jalur ekstraksi fitur. MobileNet digunakan untuk mengekstraksi fitur lokal seperti tekstur, tepi, dan detail lesi, sedangkan *Vision Transformer* memproses citra sebagai sekumpulan

patch untuk menangkap hubungan global dan konteks spasial.

Tahap *feature-level fusion* dilakukan dengan menggabungkan fitur dari kedua model menjadi satu representasi terpadu, sebagaimana dirumuskan pada Rumus 2.7.

$$F_{fused} = \text{concat}(F_{\text{MobileNet}}, F_{\text{ViT}}) \tag{2.7}$$

Pada penelitian acuan, fitur gabungan kemudian dilakukan *dimension reduction* menggunakan *Principal Component Analysis* (PCA) sebelum akhirnya dilanjutkan dengan klasifikasi menggunakan *bagging-based logistic regression*. Struktur arsitektur *hybrid* tersebut ditunjukkan pada Gambar 2.4.



Gambar 2.4. Arsitektur *hybrid* MobileNet–*Vision Transformer* pada penelitian acuan

Penelitian ini mengadopsi konsep *dual-stream* dan *feature-level fusion* dari arsitektur acuan tersebut, namun dengan beberapa perbedaan mendasar. MobileNetV2 digunakan sebagai pengganti MobileNet versi pertama karena arsitekturnya yang lebih efisien dengan *inverted residual block*. Reduksi dimensi menggunakan PCA tidak diterapkan karena fitur gabungan langsung diteruskan ke *MLP classification head* dengan struktur 2048→512→256→3, menggantikan pendekatan *bagging-based logistic regression* pada penelitian acuan.

2.8 *Transfer Learning* dan *Fine-Tuning*

Transfer learning adalah pendekatan pelatihan model di mana bobot awal model tidak dimulai dari kondisi acak, melainkan diinisialisasi menggunakan bobot yang telah dioptimalkan melalui proses pelatihan sebelumnya pada dataset skala besar. Bobot awal ini disebut bobot *pretrained*. Pendekatan ini memungkinkan

model untuk langsung memiliki kemampuan representasi fitur visual dasar sejak awal pelatihan, sehingga proses adaptasi terhadap dataset target menjadi lebih efisien, terutama pada kondisi di mana data berlabel tersedia dalam jumlah terbatas [34].

Fine-tuning adalah tahap lanjutan dari *transfer learning*, di mana sebagian atau seluruh bobot model *pretrained* disesuaikan kembali terhadap data target melalui proses pelatihan. Berbeda dengan pendekatan *feature extraction* yang membekukan seluruh bobot *backbone* dan hanya melatih *classification head*, *fine-tuning* membuka sejumlah layer untuk dilatih agar representasi fitur yang dihasilkan lebih sesuai dengan karakteristik domain data target.

Salah satu strategi *fine-tuning* yang umum digunakan adalah *layer-wise unfreezing*, yaitu membuka layer secara bertahap dari layer terdalam menuju layer terluar. Pendekatan ini digunakan karena layer yang lebih dalam pada model *pretrained* umumnya menyimpan representasi fitur yang lebih umum dan dapat ditransfer ke domain baru, sedangkan layer yang lebih dangkal menyimpan fitur yang lebih spesifik terhadap domain asal. Dengan membuka layer secara bertahap, model dapat beradaptasi terhadap domain baru tanpa langsung mengubah seluruh bobot *pretrained* secara agresif, sehingga risiko *catastrophic forgetting* dapat diminimalkan [35]. Pada arsitektur *hybrid* yang menggabungkan lebih dari satu *backbone*, strategi pembekuan bertahap ini dapat diterapkan secara berbeda untuk masing-masing *backbone* sesuai dengan ukuran parameter dan karakteristik representasi *pretrained*-nya. *Backbone* dengan jumlah parameter yang lebih sedikit dapat dilatih secara penuh sejak awal proses pelatihan, sedangkan *backbone* dengan representasi *pretrained* yang lebih kompleks memerlukan strategi pembekuan bertahap agar adaptasinya terhadap domain target berlangsung secara lebih terkendali.

2.8.1 *Learning Rate Scheduler*: *Linear Warmup* dan *Cosine Annealing*

Jadwal *learning rate* digunakan untuk mengatur nilai *learning rate* secara dinamis selama proses pelatihan. Pada penelitian ini digunakan kombinasi dua strategi: *linear warmup* pada fase awal, diikuti *cosine annealing* pada fase selanjutnya.

Linear warmup menaikkan *learning rate* secara linier dari nilai awal yang kecil menuju nilai target selama sejumlah epoch awal. Strategi ini diperlukan karena pada awal pelatihan, *classification head* masih memiliki bobot acak sehingga

gradiennya tidak stabil. Dengan menaikkan *learning rate* secara bertahap, model dapat menstabilkan parameternya terlebih dahulu sebelum pelatihan penuh dimulai, sehingga bobot *pretrained* dari *backbone* tidak langsung terganggu oleh gradien yang besar [36].

Setelah fase *warmup*, *cosine annealing* digunakan untuk menurunkan *learning rate* secara halus mengikuti fungsi kosinus dari nilai puncak menuju nilai minimum. Nilai *learning rate* pada *cosine annealing* dirumuskan sebagai berikut [37]:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right) \tag{2.8}$$

Keterangan:

- η_t merupakan *learning rate* pada epoch ke-t.
- $\eta_{\max}$ merupakan *learning rate* maksimum.
- $\eta_{\min}$ merupakan *learning rate* minimum.
- t merupakan epoch saat ini dihitung dari akhir fase *warmup*.
- T merupakan total epoch pada fase *cosine annealing*.

2.9 *Contrast Limited Adaptive Histogram Equalization* (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) merupakan teknik peningkatan kontras citra yang bekerja secara adaptif pada wilayah lokal. Berbeda dengan perataan histogram global yang memproses seluruh citra sekaligus, CLAHE membagi citra menjadi sejumlah wilayah kecil yang disebut *tile*, kemudian menerapkan perataan histogram secara independen pada setiap wilayah tersebut. Pendekatan lokal ini memungkinkan peningkatan kontras yang lebih merata pada seluruh bagian citra, termasuk pada area yang gelap maupun terang secara bersamaan [38].

Untuk mencegah amplifikasi *noise* yang berlebihan, CLAHE membatasi nilai penguatan kontras pada setiap *tile* menggunakan mekanisme *clip limit*. Proses ini dapat dirumuskan sebagai:

$$\mathrm{CLAHE}(I) = \mathrm{Clip}\left(\mathrm{AHE}(I), \beta\right) \tag{2.9}$$

Keterangan:

- I merupakan citra input.
- AHE(I) merupakan hasil *Adaptive Histogram Equalization* pada citra I.
- β merupakan nilai *clip limit* yang membatasi amplifikasi kontras maksimum.

Penerapan CLAHE pada citra *ultrasound* payudara didasarkan pada pertimbangan bahwa citra *ultrasound* umumnya memiliki distribusi intensitas yang tidak merata akibat *speckle noise* dan variasi *echogenicity* pada jaringan. Penelitian sebelumnya menunjukkan bahwa CLAHE efektif meningkatkan kualitas tepi dan mempertajam batas lesi pada citra *ultrasound* tanpa memperkenalkan amplifikasi *noise* yang berlebihan [39].

2.10 *Stratified K-Fold Cross-Validation*

Cross-validation adalah metode evaluasi model yang dilakukan dengan membagi dataset menjadi beberapa subset untuk memperoleh estimasi performa yang lebih stabil dan tidak hanya bergantung pada satu pembagian data. Pada metode *K-fold cross-validation*, dataset dibagi menjadi K bagian atau *fold*. Pada setiap iterasi, satu dari K *fold* digunakan sebagai data *validation*, sedangkan $K-1$ *fold* lainnya digunakan sebagai data *training*. Proses ini dilakukan sebanyak K kali sehingga setiap *fold* memperoleh giliran sebagai data *validation* tepat satu kali [40].

Stratified K-fold cross-validation merupakan varian dari *K-fold* yang menjaga proporsi distribusi kelas pada setiap *fold* agar tetap sesuai dengan distribusi kelas pada dataset keseluruhan. Stratifikasi ini penting pada dataset dengan distribusi kelas tidak seimbang, karena tanpa stratifikasi beberapa *fold* dapat memiliki komposisi kelas yang berbeda jauh dari dataset asli dan menghasilkan estimasi performa yang kurang representatif [41].

Hasil evaluasi dari seluruh *fold* dirata-ratakan untuk memperoleh estimasi performa keseluruhan. Nilai *mean* menunjukkan performa rata-rata model, sedangkan *standard deviation* menunjukkan kestabilan performa antar-*fold*.

2.11 Penanganan Ketidakseimbangan Kelas

Ketidakseimbangan kelas (*class imbalance*) terjadi ketika jumlah sampel pada setiap kelas dalam dataset tidak seimbang. Pada kondisi ini, model yang

dilatih tanpa penanganan khusus cenderung bias terhadap kelas mayoritas, karena model dapat mencapai akurasi tinggi dengan memprediksi kelas mayoritas secara konsisten tanpa benar-benar mengenali karakteristik kelas minoritas [41].

Salah satu pendekatan untuk mengatasi ketidakseimbangan kelas adalah *WeightedRandomSampler*, yaitu metode pengambilan sampel yang memberikan bobot berbeda pada setiap sampel selama proses pelatihan. Sampel dari kelas yang jumlahnya lebih sedikit mendapatkan bobot lebih besar sehingga probabilitas terpilihnya lebih tinggi, sedangkan sampel dari kelas mayoritas mendapatkan bobot lebih kecil. Dengan demikian, distribusi kelas pada setiap *mini-batch* menjadi lebih seimbang tanpa perlu menambah atau menghapus data dari dataset asli.

Bobot untuk setiap sampel dihitung menggunakan rumus berikut:

$$w_i = \frac{1}{n_c} \tag{2.10}$$

Keterangan:

- w_i merupakan bobot sampel ke-i.
- n_c merupakan jumlah sampel pada kelas c yang merupakan kelas dari sampel ke-i.

Pendekatan ini lebih populer dibandingkan *oversampling* sederhana karena tidak menghasilkan duplikasi data secara eksplisit, sehingga tidak menambah risiko *overfitting* akibat pengulangan sampel yang identik [41].

2.12 Evaluasi Model Klasifikasi

Evaluasi performa model klasifikasi dilakukan untuk mengetahui seberapa baik model dalam menghasilkan prediksi yang tepat dan konsisten. Salah satu pendekatan yang umum digunakan adalah *confusion matrix*, yaitu tabel yang menunjukkan hubungan antara label hasil prediksi dan label sebenarnya. Melalui tabel ini, kinerja model dapat dianalisis berdasarkan empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Dari keempat komponen tersebut, beberapa metrik evaluasi dapat dihitung untuk menilai performa model. Salah satunya adalah *accuracy*, yang menunjukkan perbandingan antara jumlah prediksi yang benar dengan keseluruhan data yang diuji, sebagaimana ditampilkan pada Rumus 2.11.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.11}$$

Meskipun *accuracy* mudah dipahami, metrik ini kurang representatif pada kasus dengan distribusi data yang tidak seimbang. Oleh karena itu, metrik *precision* dan *recall* juga digunakan, sebagaimana ditunjukkan pada Rumus 2.12 dan Rumus 2.13.

$$Precision = \frac{TP}{TP + FP} \tag{2.12}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.13}$$

Selain itu, digunakan pula *F1-score* sebagai rata-rata harmonik dari *precision* dan *recall*, sebagaimana ditunjukkan pada Rumus 2.14.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.14}$$

Penggunaan kombinasi metrik ini memungkinkan evaluasi performa model secara lebih menyeluruh, terutama dalam kasus klasifikasi medis seperti deteksi kanker payudara, di mana kesalahan dalam mendeteksi kasus positif (*false negative*) dapat berdampak serius.

Selain metrik berbasis prediksi kelas, penelitian ini juga menggunakan *Intersection over Union* (IoU) untuk mengevaluasi kesesuaian antara area fokus model pada *heatmap* Grad-CAM++ dengan lokasi lesi yang sebenarnya. IoU dihitung sebagai rasio antara area irisan dan area gabungan dari dua peta biner, sebagaimana dirumuskan pada Persamaan 2.15.

$$IoU = \frac{A \cap B}{A \cup B} \tag{2.15}$$

Keterangan:

- A merupakan area aktif hasil binarisasi *heatmap*.
- B merupakan area lesi berdasarkan *ground-truth mask*.

- $A \cap B$ merupakan area irisan antara keduanya.
- $A \cup B$ merupakan area gabungan antara keduanya.

Semakin tinggi nilai IoU, semakin besar kesesuaian antara area yang diperhatikan model dan lokasi lesi yang sebenarnya secara klinis.

2.13 *Explainable Artificial Intelligence* (XAI)

Explainable Artificial Intelligence (XAI) merupakan pendekatan dalam kecerdasan buatan yang digunakan untuk meningkatkan transparansi dan interpretabilitas model, terutama pada model *black-box* seperti *deep learning*. Model *deep learning* sering kali sulit dipahami karena arsitekturnya yang kompleks, sehingga diperlukan metode tambahan untuk menjelaskan proses pengambilan keputusan model [42]. Dalam bidang medis, interpretabilitas menjadi aspek krusial karena keputusan sistem AI dapat memengaruhi proses diagnosis dan penanganan pasien.

Pada analisis citra medis, XAI digunakan untuk menunjukkan bagian citra yang berpengaruh terhadap keputusan model. Dengan demikian, tenaga medis seperti radiolog dapat memeriksa apakah model berfokus pada area yang relevan secara klinis [43]. Salah satu metode XAI yang digunakan dalam penelitian ini adalah Grad-CAM++, yang akan dibahas pada subbab berikutnya.

2.14 Grad-CAM++

Grad-CAM++ merupakan pengembangan dari metode *Gradient-weighted Class Activation Mapping* (Grad-CAM) yang diusulkan untuk mengatasi keterbatasan Grad-CAM dalam melokalisasi objek kecil. Pada Grad-CAM standar, bobot kepentingan setiap *channel* dihitung menggunakan *global average pooling* sederhana terhadap gradien, sehingga seluruh piksel pada *feature map* berkontribusi secara merata. Pendekatan ini menghasilkan *heatmap* yang cenderung kasar dan kurang terlokalisasi pada area kecil [44].

Grad-CAM++ memperbaiki hal tersebut dengan menghitung bobot secara per-piksel menggunakan turunan orde kedua dan ketiga dari gradien, sehingga area kecil yang memberikan kontribusi tinggi terhadap prediksi mendapat perhatian yang lebih presisi [17]. Bobot per piksel pada Grad-CAM++ dihitung menggunakan Rumus 2.16.

Pemilihan Grad-CAM++ pada penelitian ini didasarkan pada kebutuhan lokalisasi yang lebih spesifik terhadap area lesi. Berbeda dengan Grad-CAM yang memberikan satu bobot global untuk setiap *feature map*, Grad-CAM++ menghitung koefisien kepentingan secara spasial menggunakan turunan orde tinggi. Mekanisme tersebut memungkinkan bagian-bagian kecil yang memiliki kontribusi kuat terhadap prediksi tetap terwakili dalam peta aktivasi [17]. Pada penelitian acuan, Grad-CAM menghasilkan representasi atensi yang lebih luas, sedangkan Grad-CAM++ menghasilkan peta yang lebih detail dalam memperlihatkan batas lesi dan perubahan visual yang subtil [18]. Karakteristik tersebut sesuai dengan citra *ultrasound* payudara yang memiliki *speckle noise*, kontras rendah, serta batas lesi yang tidak selalu tegas. Oleh karena itu, Grad-CAM++ dinilai lebih sesuai dengan tujuan penelitian ini, yaitu mengevaluasi apakah perhatian spasial model beririsan dengan *ground-truth mask* lesi, bukan hanya menghasilkan penjelasan visual secara umum.

$$\alpha_{kij}^{c} = \frac{\frac{\partial^2 y^c}{\partial (A_{ij}^k)^2}}{2\frac{\partial^2 y^c}{\partial (A_{ij}^k)^2} + \sum_{a,b} A_{ab}^k \frac{\partial^3 y^c}{\partial (A_{ij}^k)^3}} \tag{2.16}$$

Bobot untuk setiap *channel* kemudian dihitung dengan menggabungkan bobot piksel dan gradien positif, sebagaimana ditunjukkan pada Rumus 2.17.

$$w_k^c = \sum_i \sum_j \alpha_{kij}^c \cdot \text{ReLU}\left(\frac{\partial y^c}{\partial A_{ij}^k}\right) \tag{2.17}$$

Peta aktivasi akhir dihasilkan dengan menggabungkan *feature map* menggunakan bobot tersebut dan menerapkan fungsi ReLU, sebagaimana ditunjukkan pada Rumus 2.18.

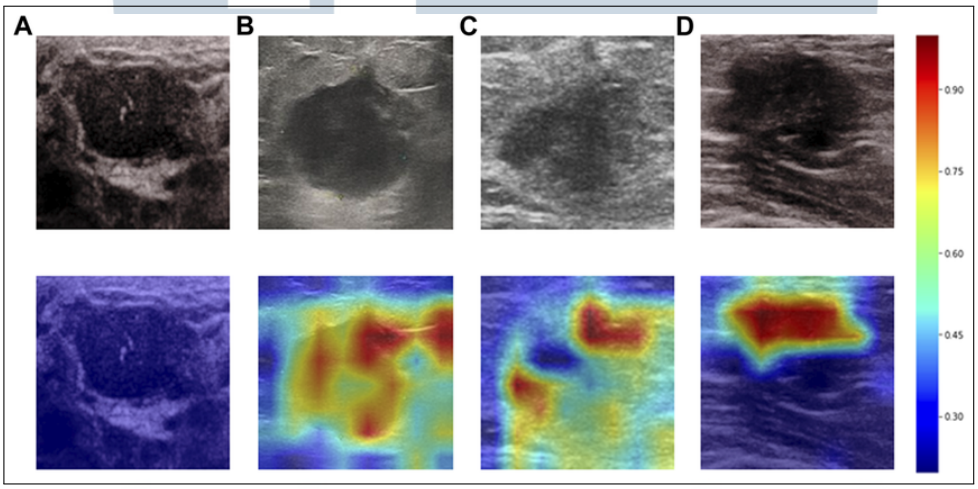
$$L_{\text{Grad-CAM++}}^c = \text{ReLU}\left(\sum_k w_k^c A^k\right) \tag{2.18}$$

Keterangan:

- α_{kij}^c merupakan bobot piksel (i, j) pada *channel* ke-k untuk kelas c.
- A^k merupakan *feature map* pada *channel* ke-k.

- y^c merupakan skor prediksi untuk kelas c.
- w_k^c merupakan bobot gabungan untuk *channel* ke-k.
- $L^c_{\text{Grad-CAM++}}$ merupakan peta aktivasi akhir untuk kelas c.

Hasil akhir berupa *heatmap* yang dapat di-*overlay* dengan citra asli untuk memberikan visualisasi yang mudah dipahami, sebagaimana ditunjukkan pada Gambar 2.5. Penggunaan bobot berbasis gradien orde tinggi menjadikan Grad-CAM++ lebih sesuai untuk analisis lesi payudara pada citra *ultrasound*, di mana lesi sering kali berukuran kecil dan memiliki batas yang tidak tegas.



Gambar 2.5. Contoh hasil visualisasi Grad-CAM++

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA