

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker payudara adalah salah satu penyebab utama kematian pada perempuan di seluruh dunia. Diperkirakan terdapat sekitar 2,3 juta kasus baru kanker payudara di seluruh dunia pada tahun 2020 [1]. Tingkat mortalitas yang signifikan menjadikan deteksi dini sebagai faktor krusial dalam meningkatkan peluang kesembuhan pasien [2]. Semakin cepat kanker teridentifikasi, semakin besar kemungkinan terapi dapat dilakukan secara tepat sasaran dan efektif. Oleh karena itu, sistem diagnosis berbantuan komputer (*Computer-Aided Diagnosis/CAD*) yang akurat, andal, dan dapat diinterpretasikan dikembangkan dan menjadi kebutuhan yang semakin mendesak dalam bidang kesehatan [3, 4].

Dalam praktik klinis, berbagai modalitas pencitraan medis digunakan untuk mengevaluasi kelainan pada jaringan payudara, seperti mammografi, *Magnetic Resonance Imaging* (MRI), dan *ultrasound*. *Ultrasound* memiliki peran penting dalam pemeriksaan payudara karena bersifat non-invasif, tidak menggunakan radiasi ionisasi, relatif terjangkau, serta mampu memberikan gambaran jaringan secara *real-time*. Modalitas ini banyak digunakan untuk mengevaluasi kelainan yang teraba, mengkarakterisasi massa, membedakan lesi kistik dan padat, serta menilai temuan yang memerlukan pemeriksaan lebih lanjut [5, 6].

Penggunaan *ultrasound* dalam pemeriksaan payudara tidak terbatas pada satu kelompok usia atau kondisi jaringan tertentu. Dalam praktik klinis, pemilihan modalitas pencitraan disesuaikan dengan usia, gejala, karakteristik jaringan payudara, serta hasil pemeriksaan sebelumnya. Pada kondisi tertentu, *ultrasound* dapat digunakan sebagai pemeriksaan awal, sedangkan pada kondisi lainnya digunakan sebagai pemeriksaan tambahan untuk memperoleh informasi yang lebih rinci mengenai karakteristik suatu lesi. Variasi karakteristik jaringan antarindividu, termasuk kepadatan payudara, juga menyebabkan relevansi penggunaan *ultrasound* tidak dapat ditentukan hanya berdasarkan usia pasien [7, 8].

Meskipun memiliki berbagai keunggulan, citra *ultrasound* payudara memiliki karakteristik visual yang kompleks, seperti *speckle noise* yang tinggi, kontras yang rendah, variasi intensitas jaringan, serta batas lesi yang sering kali tidak tegas. Kondisi tersebut dapat menyulitkan proses identifikasi dan karakterisasi lesi, serta

menyebabkan interpretasi citra bergantung secara signifikan pada pengalaman dan keahlian radiolog. Oleh karena itu, dibutuhkan pendekatan komputasional yang mampu membantu mengekstraksi pola visual dari citra *ultrasound* secara lebih konsisten dan objektif.

Perkembangan teknologi kecerdasan buatan, khususnya *deep learning*, telah membuka peluang besar dalam analisis citra medis secara otomatis. *Convolutional Neural Network* (CNN) telah terbukti sangat baik dalam melakukan ekstraksi fitur lokal seperti tekstur, tepi, dan pola mikro pada citra [9]. Keunggulan CNN terletak pada mekanisme konvolusi yang secara hierarkis mengekstraksi fitur dari level rendah hingga tinggi. Namun, CNN memiliki keterbatasan dalam memodelkan hubungan global antar bagian citra [10]. Operasi konvolusi bekerja secara lokal dengan *receptive field* yang terbatas, sehingga pemodelan konteks global biasanya diperoleh secara bertahap melalui penumpukan lapisan yang lebih dalam [11]. Dalam konteks analisis medis, informasi global seperti distribusi pola jaringan di sekitar lesi dan bentuk keseluruhan massa dapat memiliki peran penting dalam pengambilan keputusan diagnostik.

Keterbatasan ini mendorong pengembangan arsitektur yang memiliki kemampuan untuk menangkap hubungan global antar bagian citra dengan lebih efektif. *Vision Transformer* (ViT) awalnya diperkenalkan sebagai arsitektur yang memiliki mekanisme berbasis *self-attention* yang mampu memodelkan hubungan global antar bagian citra secara langsung [12]. Berbeda dengan CNN yang menggunakan filter konvolusi, ViT membagi citra menjadi *patch-patch* kecil dan memprosesnya sebagai urutan token, sehingga setiap bagian citra dapat berinteraksi dengan bagian lainnya melalui mekanisme *attention*. Pendekatan ini meningkatkan kemungkinan model untuk menangkap konteks bersifat global secara lebih eksplisit.

Sebuah penelitian menunjukkan bahwa arsitektur *hybrid* yang melakukan penggabungan CNN dan *Vision Transformer* berpotensi memanfaatkan keunggulan masing-masing pendekatan, yaitu kemampuan CNN untuk mengekstraksi fitur yang bersifat lokal serta kemampuan *Vision Transformer* dalam menangkap konteks global [13]. Kombinasi keduanya dinilai dapat menghasilkan representasi fitur yang lebih menyeluruh dibandingkan penggunaan satu arsitektur secara terpisah. Meskipun demikian, penerapan arsitektur *hybrid* tersebut masih belum banyak dikaji pada citra *ultrasound* payudara. Hal ini penting karena citra *ultrasound* memiliki karakteristik visual yang berbeda dari mammogram, seperti tingkat *noise* yang lebih tinggi dan batas lesi yang cenderung kurang tegas. Oleh karena itu,

masih terdapat ruang penelitian untuk mengevaluasi adaptasi arsitektur *hybrid CNN* dan *Vision Transformer* pada klasifikasi citra *ultrasound* payudara.

Selain aspek performa klasifikasi, tantangan lain dalam implementasi *deep learning* pada bidang medis adalah masalah interpretabilitas. Pada umumnya, model *deep learning* memiliki sifat *black-box*, sehingga proses pengambilan keputusan tidak dapat dijelaskan secara transparan [14]. Dalam konteks klinis, interpretabilitas menjadi faktor penting karena tenaga medis perlu memahami dasar pertimbangan model sebelum mengintegrasikannya dalam proses diagnosis [15]. Tanpa interpretasi yang memadai, model berisiko memanfaatkan artefak citra atau pola yang tidak relevan secara klinis.

Salah satu metode *Explainable Artificial Intelligence* yang banyak digunakan dalam analisis citra adalah *Gradient-weighted Class Activation Mapping* (Grad-CAM). Metode ini menghasilkan peta aktivasi kelas yang memperlihatkan bagian citra yang berkontribusi terhadap prediksi model. Namun, Grad-CAM menghitung bobot kepentingan setiap *feature map* melalui rata-rata global gradien, sehingga peta aktivasi yang dihasilkan cenderung bersifat kasar dan dapat kehilangan informasi spasial pada objek berukuran kecil [16, 17]. Keterbatasan tersebut menjadi penting pada citra *ultrasound* payudara karena lesi dapat berukuran kecil, memiliki kontras rendah, dan mempunyai batas yang tidak tegas. Pada kondisi tersebut, peta aktivasi yang terlalu luas belum cukup untuk memastikan bahwa keputusan model benar-benar didasarkan pada area lesi dan bukan pada jaringan sekitar atau artefak citra.

Grad-CAM++ dikembangkan untuk menghasilkan lokalisasi yang lebih rinci melalui pembobotan gradien secara spasial sehingga piksel yang memberikan kontribusi tinggi terhadap prediksi memperoleh bobot yang lebih spesifik [17]. Penelitian acuan menerapkan Grad-CAM dan Grad-CAM++ serta menunjukkan bahwa Grad-CAM++ menghasilkan peta atensi yang lebih detail dan presisi dalam memperlihatkan batas lesi dan perubahan visual yang subtil [18]. Oleh karena itu, penelitian ini memilih Grad-CAM++ sebagai metode interpretabilitas utama. Pemilihan tersebut tidak hanya bertujuan menghasilkan visualisasi, tetapi juga memungkinkan pengujian apakah area perhatian model sesuai dengan anotasi lesi melalui evaluasi kuantitatif menggunakan *Intersection over Union* (IoU). Evaluasi ini penting karena tingkat akurasi klasifikasi yang tinggi belum menjamin bahwa model mengambil keputusan berdasarkan area yang relevan secara klinis.

Secara keseluruhan, penelitian ini menggunakan arsitektur dua cabang paralel yang memproses citra *ultrasound* menggunakan MobileNetV2 dan *Vision*

Transformer. MobileNetV2 digunakan untuk mengekstraksi fitur lokal seperti tekstur, tepi, dan karakteristik visual lesi, sedangkan *Vision Transformer* digunakan untuk menangkap hubungan global dan konteks antarbagian citra. Fitur yang dihasilkan oleh kedua cabang kemudian digabungkan melalui *feature-level fusion* dan diteruskan ke *classification head* untuk menghasilkan prediksi kelas *benign*, *malignant*, atau *normal*. Setelah proses klasifikasi, Grad-CAM++ diterapkan pada cabang MobileNetV2 untuk menghasilkan *heatmap* yang menunjukkan area citra yang berkontribusi terhadap keputusan model. Kesesuaian area perhatian tersebut selanjutnya dievaluasi dengan membandingkan *heatmap* Grad-CAM++ terhadap *ground-truth mask* lesi menggunakan metrik *Intersection over Union* (IoU).

Berdasarkan uraian tersebut, penelitian ini mengembangkan arsitektur *hybrid* MobileNetV2 dan *Vision Transformer* untuk klasifikasi lesi payudara berbasis citra *ultrasound* pada dataset BUSI yang terdiri dari 780 citra dengan tiga kelas, yaitu *benign*, *malignant*, dan *normal*. Selain mengevaluasi performa klasifikasi menggunakan *stratified 5-fold cross-validation*, penelitian ini juga menganalisis interpretabilitas model menggunakan Grad-CAM++ dan mengevaluasinya secara kuantitatif menggunakan metrik *Intersection over Union* (IoU) terhadap *ground-truth mask* lesi. Dengan demikian, penelitian ini diharapkan mampu memberikan kontribusi yang positif dalam proses pengembangan sistem diagnosis berbasis *deep learning* yang akurat dan juga transparan serta dapat dipertanggungjawabkan secara klinis.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan arsitektur *hybrid* MobileNetV2 dan *Vision Transformer* untuk klasifikasi lesi payudara pada citra *ultrasound* berdasarkan dataset BUSI?
2. Seberapa tinggi tingkat kinerja model *hybrid* MobileNetV2 dan *Vision Transformer* dalam mengklasifikasikan lesi payudara pada citra *ultrasound* berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*?
3. Bagaimana kesesuaian area atensi yang dihasilkan oleh Grad-CAM++ dengan lokasi lesi aktual pada citra *ultrasound* berdasarkan evaluasi kuantitatif menggunakan metrik *Intersection over Union* (IoU)?

1.3 Batasan Permasalahan

Dengan tujuan menjaga fokus penelitian ini pada tujuan yang telah ditetapkan, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan adalah *Breast Ultrasound Images Dataset* (BUSI) yang diperkenalkan dan didokumentasikan oleh Al-Dhabyani *et al.* melalui artikel data pada jurnal *Data in Brief* [19].
2. Penelitian ini berfokus pada klasifikasi tiga kelas, yaitu *benign*, *malignant*, dan *normal*.
3. Analisis interpretabilitas model dilakukan pada layer `features[-1]` dari cabang MobileNetV2.
4. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* dengan skema *stratified 5-fold cross-validation*.
5. Penelitian ini tidak membahas aspek implementasi klinis secara langsung, melainkan berfokus pada analisis performa dan interpretabilitas model dalam konteks penelitian.

1.4 Tujuan Penelitian

Penelitian memiliki tujuan sebagai berikut:

1. Mengimplementasikan arsitektur *hybrid* MobileNetV2 dan *Vision Transformer* untuk klasifikasi lesi payudara pada citra *ultrasound* menggunakan dataset BUSI.
2. Mengevaluasi tingkat kinerja model *hybrid* MobileNetV2 dan *Vision Transformer* dalam mengklasifikasikan lesi payudara pada citra *ultrasound* berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*.
3. Menganalisis kesesuaian area atensi yang dihasilkan oleh Grad-CAM++ dengan lokasi lesi aktual pada citra *ultrasound* secara kuantitatif menggunakan metrik *Intersection over Union* (IoU).

1.5 Manfaat Penelitian

Manfaat dari penelitian ini diharapkan sebagai berikut:

1. Berkontribusi dalam pengembangan klasifikasi citra medis dengan *deep learning*, khususnya dalam pemanfaatan arsitektur *hybrid* MobileNetV2 dan *Vision Transformer* untuk analisis citra *ultrasound* payudara.
2. Menyediakan analisis interpretabilitas model menggunakan Grad-CAM++ sehingga dapat membantu memahami area citra yang berpengaruh terhadap keputusan model, sekaligus mengevaluasinya secara kuantitatif menggunakan IoU.
3. Mendukung peningkatan transparansi dan kepercayaan terhadap model *deep learning* dalam analisis citra medis melalui pemanfaatan *Explainable Artificial Intelligence* (XAI) berbasis Grad-CAM++.
4. Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan penerapan *deep learning* dan *Explainable AI* dalam bidang analisis citra medis, khususnya untuk deteksi dan klasifikasi kanker payudara.

1.6 Sistematika Penulisan

Berikut adalah uraian singkat mengenai struktur isi penulisan laporan penelitian ini:

- **Bab 1 PENDAHULUAN**

Bab ini membahas latar belakang permasalahan, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- **Bab 2 LANDASAN TEORI**

Bab ini membahas landasan teori yang menjadi dasar penelitian, meliputi kanker payudara, citra *ultrasound* payudara, *deep learning*, CNN, MobileNet, *Vision Transformer*, arsitektur *hybrid*, *transfer learning*, CLAHE, *stratified k-fold cross-validation*, penanganan ketidakseimbangan kelas, metrik evaluasi, XAI, dan Grad-CAM++.

- **Bab 3 METODOLOGI PENELITIAN**

Bab ini menjelaskan metodologi penelitian yang mencakup alur penelitian,

dataset BUSI, tahap *data collecting* dan *data cleaning*, pembagian data, *preprocessing*, perancangan model *hybrid*, strategi *transfer learning*, dan skema evaluasi model.

- **Bab 4 HASIL DAN DISKUSI**

Bab ini menyajikan hasil implementasi dan pengujian model, meliputi hasil pelatihan pada setiap *fold*, analisis *loss* dan akurasi, *confusion matrix*, serta analisis interpretabilitas menggunakan Grad-CAM++.

- **Bab 5 KESIMPULAN DAN SARAN**

Bab ini berisi simpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.

