

# Pendekatan Ensemble Multi-Arsitektur Convolutional Neural Network melalui Soft Voting untuk Klasifikasi Citra Histopatologi Kanker Payudara

Shelomita Fitriyani\*, Sindhu Rakasiwi

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: <sup>1,\*</sup>111202214319@mhs.dinus.ac.id, <sup>2</sup>sindhu.rakasiwi@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202214319@mhs.dinus.ac.id

Submitted: 27/11/2025; Accepted: 26/12/2025; Published: 26/12/2025

**Abstrak**—Kanker payudara merupakan salah satu penyebab kematian tertinggi pada perempuan sehingga diperlukan metode deteksi citra medis yang akurat, konsisten, dan mampu menangani variasi morfologis jaringan. Penelitian ini bertujuan untuk meningkatkan stabilitas dan akurasi klasifikasi citra histopatologi kanker payudara melalui pendekatan *ensemble* multiarsitektur *Convolutional Neural Network*. Dataset yang digunakan adalah *BreakHis* yang terdiri dari empat tingkat pembesaran, yaitu 40×, 100×, 200×, dan 400×. Tiga arsitektur digunakan sebagai model dasar, yaitu *VGG19*, *ResNet50*, dan *EfficientNetB0*. Seluruh citra melalui proses *preprocessing* berupa *resize* ke 224×224 piksel, normalisasi intensitas, dan *augmentasi data*. Setiap model dilatih secara independen, kemudian *output probabilitas* ketiganya digabungkan menggunakan teknik *soft voting* untuk menghasilkan prediksi akhir yang lebih stabil. Hasil eksperimen menunjukkan bahwa metode *ensemble* memberikan performa terbaik dan paling konsisten di seluruh tingkat pembesaran. Pada pembesaran 40×, model mencapai akurasi sebesar 92,00%, dengan *recall* 99,03% dan *F1-score* 94,44%. Pada pembesaran 100×, akurasi meningkat menjadi 94,56%, dengan *recall* 99,07% dan *F1-score* 96,18%. Pembesaran 200× menghasilkan akurasi 94,03%, disertai *recall* 97,61% dan *F1-score* 95,77%. Sementara itu, pada pembesaran 400×, akurasi tetap kompetitif sebesar 90,11%, dengan *recall* 95,14% dan *F1-score* 92,88%. Tingginya nilai *recall* pada seluruh pembesaran menunjukkan kemampuan model dalam mendeteksi citra *malignant* secara konsisten, sedangkan nilai *F1-score* yang stabil mengindikasikan keseimbangan yang baik antara *precision* dan *recall*. Temuan ini menegaskan bahwa penggabungan beberapa arsitektur *CNN* mampu menangkap fitur histopatologi secara lebih komprehensif sehingga berpotensi kuat digunakan sebagai sistem pendukung keputusan pada diagnosis kanker payudara berbasis citra medis.

**Kata Kunci:** Kanker Payudara; Histopatologi; Convolutional Neural Network; Ensemble Learning; Soft Voting

**Abstract**—Breast cancer is one of the leading causes of mortality among women, creating a strong need for diagnostic methods that are accurate, consistent, and capable of handling the morphological variations present in histopathological images. This study aims to improve the stability and accuracy of breast cancer histopathology image classification through an ensemble multi-architecture *Convolutional Neural Network* approach. The *BreakHis* dataset, which consists of four magnification levels 40×, 100×, 200×, and 400× was used in this research. Three architectures, *VGG19*, *ResNet50*, and *EfficientNetB0*, served as the base models. All images underwent *preprocessing*, including resizing to 224×224 pixels, pixel-intensity normalization, and data augmentation. Each model was trained independently, and their probability outputs were combined using a *soft voting* mechanism to generate the final predictions. The experimental results show that the ensemble method provides the most stable and superior performance across all magnification levels. At 40× magnification, the ensemble achieved an accuracy of 92.00%, recall of 99.03%, and F1-score of 94.44%. At 100× magnification, the accuracy increased to 94.56%, with a recall of 99.07% and an F1-score of 96.18%. The 200× level produced an accuracy of 94.03%, recall of 97.61%, and an F1-score of 95.77%. Meanwhile, at 400× magnification, the model reached an accuracy of 90.11%, recall of 95.14%, and an F1-score of 92.88%. These consistently high recall and F1-score values highlight the model's strong ability to detect malignant cases while maintaining balanced predictive performance. Overall, the findings demonstrate that combining multiple CNN architectures enhances feature representation and shows strong potential as a decision-support system for breast cancer diagnosis using histopathological images.

**Keywords:** Breast Cancer; Histopathology; Convolutional Neural Network; Ensemble Learning; Soft Voting

## 1. PENDAHULUAN

Kanker payudara masih menjadi salah satu tantangan kesehatan global dengan tingkat prevalensi yang terus meningkat setiap tahunnya. Laporan Global Cancer Observatory (GLOBOCAN) 2023 menunjukkan bahwa kanker payudara merupakan jenis kanker paling umum di seluruh dunia dengan lebih dari 2,3 juta kasus baru per tahun dan angka mortalitas mencapai lebih dari 600.000 jiwa, menjadikannya penyebab kematian tertinggi kedua akibat kanker pada perempuan [1]. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa risiko seorang perempuan mengalami kanker payudara sepanjang hidupnya telah meningkat hingga 1 dari 12 perempuan, terutama di negara berkembang yang masih menghadapi keterbatasan fasilitas diagnostik dan variabilitas kualitas pemeriksaan mikroskopis [2]. Analisis citra histopatologi hingga kini tetap menjadi standar emas dalam diagnosis kanker karena mampu menampilkan karakteristik jaringan secara mikroskopis, termasuk distribusi nukleus, pleomorfisme, struktur stroma, serta pola proliferasi sel [3]. Meskipun demikian, proses analisis manual sangat dipengaruhi oleh keahlian patolog, kondisi visual sampel, dan faktor kelelahan, sehingga rawan subjektivitas dan inkonsistensi penilaian [4], yang pada akhirnya dapat menurunkan akurasi diagnosis di lingkungan klinis.

Perkembangan metode kecerdasan buatan, khususnya *Convolutional Neural Network* (CNN), telah memberikan lompatan signifikan dalam analisis citra medis karena kemampuannya mengekstraksi fitur visual secara otomatis dan hierarkis tanpa memerlukan rekayasa fitur manual [5]. Sejumlah penelitian menunjukkan bahwa CNN mampu mengenali pola tekstur halus, variasi sel, serta karakteristik morfologis yang menjadi indikator utama pada

citra histopatologi kanker payudara [6]. Penerapan CNN pada dataset besar juga telah menghasilkan performa klasifikasi yang kompetitif, baik dalam tugas biner maupun multikelas [7]. Namun demikian, citra histopatologi memiliki karakteristik yang sangat beragam. Variasi warna, distribusi tekstur, artefak pewarnaan, serta perbedaan pola morfologi menyebabkan model tunggal sering kali tidak mampu menangkap keseluruhan informasi visual secara menyeluruh [8]. Selain itu, citra histopatologi umumnya disediakan dalam berbagai tingkat pembesaran yang memunculkan pola visual pada skala berbeda, sehingga memerlukan pendekatan yang mampu memahami konteks multi-skala [9]. Masalah tambahan yang sering dilaporkan adalah ketidakseimbangan jumlah sampel *benign* dan *malignant*, yang berdampak pada bias prediksi [10].

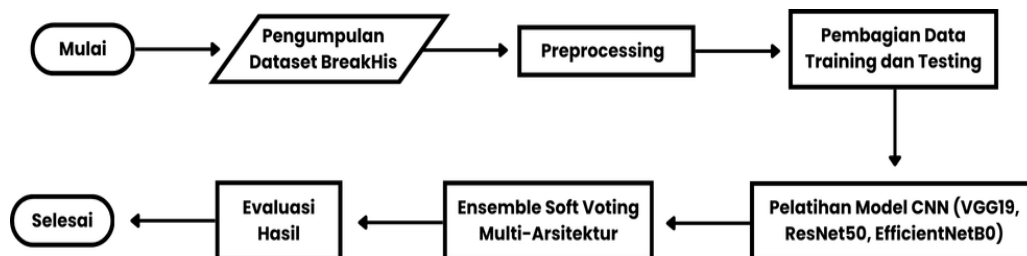
Berbagai arsitektur CNN telah dikembangkan untuk mengatasi tantangan tersebut. *VGG19* dikenal efektif dalam mengekstraksi fitur tekstur halus melalui lapisan konvolusi bertingkat dan struktur jaringan yang dalam [11]. *ResNet50* menawarkan keunggulan *residual learning* yang memungkinkan pelatihan jaringan sangat dalam tanpa mengalami hilangnya performa atau *degradasi gradien* [12]. Di sisi lain, *EfficientNetB0* menggunakan pendekatan *compound scaling* yang seimbang sehingga memberikan efisiensi komputasi yang tinggi tanpa mengorbankan kualitas representasi fitur [13]. Ketiga arsitektur ini telah digunakan secara luas dalam analisis citra histopatologi dan terbukti menghasilkan performa yang kuat dalam berbagai penelitian. Meskipun demikian, pendekatan menggunakan satu arsitektur saja masih rentan terhadap ketidakstabilan prediksi, terutama ketika citra memiliki variabilitas warna dan struktur yang tinggi.

Pendekatan *Ensemble learning* menjadi salah satu solusi yang semakin banyak digunakan untuk mengatasi ketidakstabilan tersebut. Ensemble CNN menggabungkan keluaran beberapa model sekaligus sehingga representasi fitur yang dihasilkan menjadi lebih komprehensif dan tidak bias terhadap karakteristik tertentu [14]. Penelitian terbaru menunjukkan bahwa metode ensemble dapat meningkatkan akurasi, stabilitas prediksi, dan ketahanan model terhadap variasi tekstur maupun pembesaran citra [15]. Pendekatan lain yang berkaitan, seperti deep ensemble dan model fusion, juga terbukti mampu mengurangi tingkat kesalahan, terutama pada kasus false negative yang sangat kritis dalam diagnosis kanker [16], serta menunjukkan generalisasi yang lebih kuat dibandingkan model tunggal [17]. Penggunaan ensemble adaptif bahkan menghasilkan peningkatan performa signifikan pada citra dengan pola warna tidak seragam [18], sementara pendekatan multi-arsitektur semakin populer karena mampu menangkap pola visual kompleks pada berbagai tingkat kedalaman jaringan [19]. Penelitian lain mengusulkan model hibrida dua tahap yang memisahkan proses ekstraksi fitur dan klasifikasi untuk menghasilkan representasi fitur yang lebih matang [20], dan optimalisasi ensemble menggunakan metaheuristik terbukti mempengaruhi kualitas bobot antar model sehingga berdampak pada peningkatan akurasi prediksi [21]. Berdasarkan kajian literatur tersebut, dapat disimpulkan bahwa meskipun telah banyak penelitian yang mengembangkan CNN maupun metode ensemble untuk klasifikasi citra histopatologi, sebagian besar penelitian belum melakukan evaluasi mendalam terhadap performa ensemble pada seluruh empat tingkat pembesaran dalam dataset BreakHis (40×, 100×, 200×, dan 400×). Selain itu, belum banyak penelitian yang menggabungkan tiga arsitektur berbeda dengan karakteristik komplementer *VGG19* yang unggul dalam fitur tekstur halus, *ResNet50* yang stabil untuk jaringan dalam, dan *EfficientNetB0* yang efisien dalam fitur multi-skala melalui mekanisme *soft voting*.

Oleh karena itu, penelitian ini mengusulkan pendekatan ensemble multi-arsitektur CNN yang menggabungkan *VGG19*, *ResNet50*, dan *EfficientNetB0* untuk menghasilkan model klasifikasi yang lebih stabil, akurat, dan konsisten di seluruh tingkat pembesaran. Kontribusi utama penelitian ini adalah menghadirkan sistem klasifikasi citra histopatologi kanker payudara dengan generalisasi lintas skala yang lebih baik dibandingkan model tunggal, sehingga berpotensi menjadi komponen penting dalam sistem pendukung diagnosis medis berbasis citra.

## 2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan melalui rangkaian proses sistematis yang dirancang untuk menghasilkan model klasifikasi citra histopatologi kanker payudara berbasis pendekatan ensemble multi-arsitektur CNN. Seluruh tahapan dirumuskan secara terstruktur mulai dari perancangan alur, pengolahan citra, pembangunan model, hingga tahap evaluasi untuk memastikan pendekatan yang diperoleh memiliki performa prediksi yang optimal dan stabil pada data kompleks seperti citra histopatologi. Pola kerja ini mengikuti prinsip pengolahan citra digital dan pembelajaran mendalam yang direkomendasikan pada studi-studi mutakhir di bidang bioinformatika dan analisis jaringan [16]. Alur keseluruhan proses penelitian ditunjukkan pada Gambar 1 berikut.



Gambar 1. Alur Model Penelitian

Diagram tersebut menjelaskan tahapan pemrosesan dataset BreakHis hingga proses evaluasi model. Alur dimulai dari pengumpulan dataset, dilanjutkan dengan *preprocessing* berupa *resize* dan normalisasi citra, kemudian pembagian data menjadi set pelatihan dan pengujian berdasarkan tingkat pembesaran (*magnification*). Selanjutnya, data digunakan untuk melatih tiga arsitektur CNN (*VGG19*, *ResNet50*, dan *EfficientNetB0*) dengan pendekatan *transfer learning* berbasis ekstraksi fitur. Probabilitas keluaran ketiga model kemudian digabungkan menggunakan metode ensemble *soft voting* untuk menghasilkan prediksi akhir yang lebih stabil dan robust. Tahapan ini dirancang untuk memaksimalkan kemampuan ekstraksi fitur morfologis jaringan melalui kombinasi multi-arsitektur yang dapat menangkap pola visual yang beragam.

## 2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah BreakHis (Breast Cancer Histopathological Image Dataset), sebuah himpunan data citra histopatologi yang banyak digunakan pada penelitian klasifikasi kanker payudara berbasis *deep learning*. Dataset ini diambil dari platform Kaggle, yang menyediakan versi terstruktur dan mudah diakses oleh peneliti, sehingga memudahkan proses pengunduhan dan pengelolaan data sebelum masuk ke tahap pra-pemrosesan. BreakHis berisi citra jaringan payudara yang diperoleh dari sampel klinis nyata menggunakan teknik pewarnaan mikroskopis hematoxylin-eosin (H&E). Seluruh citra dibagi ke dalam dua kelompok diagnostik, yaitu *benign* (jinak) dan *malignant* (ganas), dengan beragam variasi tekstur sel, kepadatan jaringan, dan karakteristik morfologi yang mencerminkan kondisi patologis sebenarnya.

Setiap citra disediakan dalam empat tingkat pembesaran mikroskopis, yaitu 40×, 100×, 200×, dan 400×, yang menghasilkan keragaman pola visual pada berbagai skala. Pembesaran rendah seperti 40× menampilkan gambaran struktur jaringan secara global, sementara pembesaran tinggi seperti 400× memperlihatkan detail mikroskopis seperti nukleus, sitoplasma, dan derajat proliferasi sel. Variasi tingkat pembesaran ini menjadi salah satu tantangan utama dalam klasifikasi otomatis, karena model harus mampu memahami pola makro maupun mikro secara konsisten. Setelah dataset diunduh dari Kaggle, seluruh citra disusun ulang ke dalam format biner dan dikelompokkan berdasarkan tingkat pembesaran. Pembagian data dilakukan menggunakan proporsi 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*), dengan alokasi yang seimbang pada setiap tingkat pembesaran untuk menghindari bias evaluasi. Ringkasan distribusi data keseluruhan berdasarkan proporsi *training* dan *testing* ditampilkan pada Tabel 1.

**Tabel 1.** Pembagian Dataset Keseluruhan

Subnet	Jumlah Citra	Presentase
Training	6.073	80%
Testing	1.518	20%

Selanjutnya, distribusi dataset berdasarkan tingkat pembesaran ditampilkan pada Tabel 2. Pembagian ini memperlihatkan bahwa BreakHis memiliki jumlah citra yang relatif seimbang di setiap zoom, sehingga model dapat mempelajari atribut visual dari berbagai skala tanpa mengalami bias akibat ketidakseimbangan data.

**Tabel 2.** Pembagian Dataset Pembesaran

Subnet	Training	Testing	Presentase	Total
40×	1.298	325	21.4%	1.625
100×	1.358	340	22.2%	1.687
200×	1.382	345	21.4%	1.625
400×	1.235	308	21.8%	1.654
Total	6.073	1.518	100%	7.591

Distribusi yang merata ini memastikan bahwa setiap model memiliki kesempatan belajar dari pola jaringan pada berbagai skala pembesaran, mulai dari kontur jaringan makro hingga pola tekstur mikroskopis yang lebih kompleks. Keceragaman ini juga membantu mengurangi risiko *class imbalance* yang sering terjadi pada dataset histopatologi. Selain itu, karena seluruh citra merupakan hasil pengambilan dari sampel jaringan nyata, variasi warna hasil pewarnaan, intensitas cahaya, dan struktur biologis memberikan kompleksitas yang tinggi dan sangat realistis. Hal tersebut menjadikan BreakHis tidak hanya relevan untuk pengembangan model CNN, tetapi juga ideal untuk mengevaluasi kinerja pendekatan *Ensemble learning*, terutama dalam konteks klasifikasi kanker payudara dimana konsistensi dan akurasi menjadi aspek yang sangat krusial.

## 2.2 Preprocessing

Tahapan *preprocessing* dilakukan untuk memastikan bahwa seluruh citra histopatologi berada pada format yang seragam sehingga model dapat melakukan ekstraksi fitur secara optimal. Seluruh citra terlebih dahulu dikonversi ke ukuran standar 224×224 piksel yang sesuai dengan kebutuhan arsitektur CNN modern. Seluruh citra terlebih dahulu diubah ukurannya menjadi 224×224 piksel. Proses *resizing* dinyatakan dengan:

$$I_{resized} = R(I_{original}, 224 \times 224) \quad (1)$$

di mana  $R(\cdot)$  adalah fungsi perubahan ukuran citra yang mempertahankan proporsi struktur morfologis. Selanjutnya, citra dinormalisasi dengan mengkonversi nilai piksel dari rentang 0–255 menjadi 0–1 menggunakan:

$$I_{norm}(x, y) = \frac{I(x, y)}{255} \quad (2)$$

Normalisasi ini bertujuan menstabilkan distribusi intensitas sehingga jaringan konvolusional dapat mempelajari fitur tekstur secara konsisten. Pendekatan ini selaras dengan temuan bahwa stabilisasi intensitas sangat membantu dalam meningkatkan performa CNN pada citra histopatologi [17]. Citra kemudian disusun ke dalam direktori biner yang terstruktur rapi berdasarkan kedua kelas (*benign* dan *malignant*) untuk mendukung pipeline pemanggilan data yang sistematis selama proses pelatihan.

### 2.3 Arsitektur Model CNN

Tiga arsitektur *Convolutional Neural Network* (CNN) digunakan sebagai dasar pemodelan, yaitu *VGG19*, *ResNet50*, dan *EfficientNetB0*. Ketiganya dipilih karena mewakili tiga pendekatan desain jaringan berbeda: arsitektur bertingkat klasik (VGG), pembelajaran residual (ResNet), serta scaling terintegrasi yang efisien (EfficientNet). Kombinasi ketiganya memberikan kemampuan ekstraksi fitur yang komplementer sehingga meningkatkan ketahanan prediksi ketika digabungkan dalam ensemble [18].

Ketiga model menggunakan bobot awal *ImageNet* dan hanya melatih bagian *fully connected classifier* pada lapisan akhir. Struktur *backbone* tetap dipertahankan agar proses pelatihan lebih stabil serta menghindari risiko *overfitting* yang sering terjadi pada citra histopatologi berukuran kecil.

#### a. VGG19

VGG19 merupakan arsitektur CNN klasik dengan susunan filter konvolusi  $3 \times 3$  yang ditumpuk secara berurutan. Pola desain yang dalam dan berlapis-lapis memungkinkan jaringan menangkap tekstur mikroskopis, pola granular, dan variasi struktur sel payudara yang penting untuk identifikasi *benign vs malignant* [19]. Operasi konvolusi pada *VGG19* dapat dituliskan dengan persamaan sederhana:

$$F_{out} = W * F_{in} + b \quad (3)$$

Pada persamaan tersebut,  $F_{in}$  merepresentasikan fitur masukan yang berasal dari citra histopatologi, sedangkan  $W$  merupakan kernel konvolusi yang berfungsi mengekstraksi pola lokal seperti tekstur dan struktur sel. Parameter  $b$  menyatakan bias yang ditambahkan untuk meningkatkan fleksibilitas fungsi aktivasi, sementara simbol  $*$  menunjukkan operasi konvolusi antara fitur masukan dan kernel. Melalui operasi ini, jaringan mampu mempelajari representasi fitur tingkat rendah hingga tinggi yang penting dalam membedakan jaringan *benign* dan *malignant*.

#### b. ResNet50

ResNet50 menggunakan konsep *residual learning* yang memungkinkan jaringan sangat dalam tanpa mengalami permasalahan *vanishing gradient*. Karakteristik ini sangat penting untuk citra histopatologi yang memiliki keragaman struktur seluler bertingkat, dari pola lokal mikroskopik hingga susunan yang lebih global. Blok residual disusun berdasarkan fungsi dasar berikut:

$$H(x) = F(x) + x \quad (4)$$

Dalam persamaan tersebut,  $F(x)$  merepresentasikan transformasi non-linear yang terdiri dari rangkaian operasi konvolusi, normalisasi batch, dan fungsi aktivasi ReLU. Sementara itu,  $x$  merupakan sinyal identitas yang dilewatkan secara langsung melalui mekanisme skip connection. Penambahan sinyal identitas ini memungkinkan jaringan mempelajari selisih representasi fitur antar lapisan, sehingga permasalahan *vanishing gradient* dapat diminimalkan dan proses pelatihan jaringan yang sangat dalam menjadi lebih stabil.

#### c. EfficientNetB0

EfficientNetB0 menggunakan strategi *compound scaling*, yaitu melakukan scaling terkoordinasi terhadap kedalaman (depth), lebar (width), dan resolusi (resolution) jaringan dengan satu koefisien  $\phi$  [20]. Konsep scaling dapat dinyatakan dengan rumus:

$$d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi \quad (5)$$

dengan kendala:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (6)$$

EfficientNetB0 menerapkan strategi *compound scaling* yang mengatur peningkatan kedalaman jaringan ( $d$ ), lebar jaringan ( $w$ ), dan resolusi citra ( $r$ ), secara terkoordinasi menggunakan satu koefisien skala  $\phi$ , sebagaimana ditunjukkan pada Persamaan (5) dan (6). Pada pendekatan ini, parameter  $d$ , mengontrol jumlah lapisan jaringan,  $w$  menentukan jumlah kanal pada setiap lapisan, dan  $r$

Keunikan inilah yang membuat ketiga arsitektur sangat ideal untuk dievaluasi dan dikombinasikan melalui metode *soft voting* ensemble, khususnya pada domain citra histopatologi yang memiliki high intra-class variance serta perbedaan struktur biologis antar sampel [20].

## 2.4 Transfer Learning

*Transfer learning* digunakan dalam penelitian ini untuk memanfaatkan representasi fitur yang telah dipelajari oleh model CNN besar yang sebelumnya dilatih pada dataset *ImageNet*, yaitu *ResNet50*, *EfficientNetB0*, dan *VGG19*. Pada penelitian ini, *transfer learning* diterapkan dalam bentuk *feature extraction*. Seluruh bobot dari model pretrained tersebut dibekukan (frozen), sehingga hanya lapisan klasifikasi baru pada bagian akhir model yang dilatih.

Lapisan *fully connected* asli dari masing-masing arsitektur dihapus dan digantikan dengan lapisan klasifikasi khusus sesuai jumlah kelas dalam dataset *BreakHis*. Dengan pendekatan ini, model pretrained bertindak sebagai ekstraktor fitur yang mengekstraksi pola tekstur, struktur sel, dan karakteristik histopatologi lainnya, sementara proses pembelajaran utama hanya terjadi pada *classifier* baru. Pemanfaatan pretrained CNN sebagai feature extractor telah terbukti efektif pada banyak penelitian citra medis, terutama ketika ukuran dataset relatif terbatas sehingga tidak memungkinkan pelatihan model besar dari awal [21].

## 2.5 Ensemble learning

*Ensemble learning* digunakan untuk menggabungkan keluaran dari beberapa model CNN berbeda (*ResNet50*, *EfficientNetB0*, dan *VGG19*) sehingga keputusan akhir menjadi lebih stabil, akurat, dan tahan terhadap variabilitas pola citra histopatologi. Pendekatan ini memanfaatkan keunggulan masing-masing arsitektur secara komplementer: *ResNet50* unggul dalam penangkapan fitur residual yang dalam, *EfficientNetB0* menawarkan efisiensi komputasi dengan kualitas fitur yang baik, dan *VGG19* memiliki kemampuan ekstraksi fitur hierarkis yang sangat kuat pada tekstur halus.

Pada penelitian ini, ensemble diterapkan pada level probability fusion, yaitu penggabungan probabilitas keluaran (softmax output) dari masing-masing model menggunakan strategi rata-rata tertimbang (weighted averaging). Setiap model memberikan skor probabilitas untuk kelas *benign* dan *malignant*, kemudian skor tersebut digabungkan untuk menghasilkan keputusan akhir. Pendekatan ini dapat mengurangi variansi prediksi dan mengatasi kelemahan model individu yang mungkin sensitif terhadap tingkat pembesaran tertentu atau variasi tekstur spesifik. Secara umum, formula ensemble probability fusion dapat dituliskan secara sederhana sebagai:

$$P_{ensemble}(c) = \sum_{i=1}^n w_i \cdot P_i(c) \quad (7)$$

Pada persamaan tersebut,  $P_{ensemble}(c)$  merepresentasikan probabilitas akhir untuk kelas  $c$  sedangkan  $P_i(c)$  adalah probabilitas kelas  $c$  yang dihasilkan oleh model ke- $i$ . Parameter  $w_i$  menunjukkan bobot kontribusi masing-masing model dalam ensemble, dan  $n$  menyatakan jumlah model yang digabungkan. Melalui formulasi ini, setiap model berkontribusi terhadap prediksi akhir sesuai dengan tingkat kepercayaannya. Pendekatan ensemble ini memberikan beberapa keuntungan penting dalam konteks klasifikasi citra histopatologi kanker payudara, antara lain mengurangi risiko *overfitting* karena setiap model memiliki bias dan struktur representasi fitur yang berbeda, meningkatkan generalisasi pada berbagai tingkat pembesaran (40×, 100×, 200×, 400×) berkat sensitivitas arsitektur yang berbeda terhadap detail maupun pola makroskopis, serta meningkatkan stabilitas prediksi pada citra yang mengandung artefak warna, noise, atau variasi morfologi sel yang ekstrem. Selain itu, kombinasi multi-arsitektur CNN terbukti mampu meningkatkan akurasi keseluruhan dibandingkan model tunggal pada data medis yang kompleks. Pada implementasinya, seluruh model menghasilkan probabilitas kelas pada data yang sama, kemudian probabilitas tersebut digabungkan menjadi prediksi final sehingga performa klasifikasi menjadi lebih konsisten dan robust terhadap variabilitas citra histopatologi.

## 2.6 Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan sistem dalam membedakan citra histopatologi *benign* dan *malignant* pada subset pengujian. Proses evaluasi mencakup perhitungan metrik utama klasifikasi biner, yaitu akurasi, precision, recall, F1-score, serta analisis *confusion matrix* untuk melihat pola kesalahan model. Penggunaan beberapa metrik evaluasi ini bertujuan memberikan gambaran kinerja model secara menyeluruh, tidak hanya dari tingkat ketepatan prediksi tetapi juga dari keseimbangan antara kesalahan positif dan negatif. Selain itu, analisis *confusion matrix* menjadi penting untuk mengidentifikasi kesalahan kritis seperti false negative yang berisiko tinggi dalam konteks diagnosis kanker payudara. Akurasi digunakan sebagai indikator umum performa, dihitung dengan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Precision (Presisi): Mengukur proporsi prediksi positif yang benar.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Recall (Sensitivitas atau True Positive Rate): Mengukur seberapa baik model mendeteksi kelas positif.

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

F1-Score: Rata-rata harmonis antara presisi dan recall, digunakan untuk dataset yang tidak seimbang.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

Pada perhitungan metrik evaluasi, true positive (TP) merepresentasikan jumlah citra malignant yang berhasil diklasifikasikan dengan benar oleh model, sedangkan true negative (TN) menunjukkan jumlah citra *benign* yang diprediksi secara tepat. Sementara itu, false positive (FP) menggambarkan kondisi ketika citra *benign* salah diklasifikasikan sebagai malignant, dan false negative (FN) menunjukkan citra malignant yang keliru diprediksi sebagai *benign*. Keempat komponen ini menjadi dasar utama dalam penghitungan akurasi, precision, recall, dan F1-score untuk mengevaluasi kinerja model klasifikasi. Selain metrik berbasis angka, *confusion matrix* disusun untuk memberikan gambaran langsung mengenai klasifikasi benar dan salah. Matriks ini membantu mengidentifikasi pola kesalahan yang mungkin muncul akibat variasi tekstur, warna, atau artefak pada citra histopatologi. Evaluasi dilakukan pada keempat tingkat pembesaran (40×, 100×, 200×, 400×), sehingga performa dapat diamati baik pada detail mikroskopis maupun pola yang lebih luas. Seluruh model individual dievaluasi, kemudian dibandingkan dengan hasil model ensemble *soft voting*.

Pendekatan ensemble ini menstabilkan prediksi dengan menggabungkan probabilitas dari tiga arsitektur CNN dan terbukti memberikan performa lebih konsisten. Secara keseluruhan, tahap evaluasi memastikan bahwa sistem klasifikasi mampu memberikan prediksi yang akurat, stabil, dan dapat diandalkan untuk analisis histopatologi kanker payudara berbasis citra.

### 3. HASIL DAN PEMBAHASAN

Sebelum masuk pada hasil pengujian, tahapan metode yang digunakan dalam penelitian ini perlu dijelaskan kembali secara ringkas untuk menunjukkan bagaimana setiap langkah berkontribusi langsung terhadap pencapaian performa model. Proses penyelesaian masalah dimulai dari pengumpulan dataset BreakHis, dilanjutkan dengan *preprocessing* yang meliputi resizing citra ke 224×224 piksel, normalisasi intensitas piksel ke rentang 0–1, serta *augmentasi* adaptif berupa *rotasi*, *flipping*, *zooming*, dan penyesuaian kecerahan. Tahap ini bertujuan menstandarisasi seluruh citra agar dapat diproses oleh arsitektur CNN modern dan mengurangi risiko *overfitting* akibat ketidakseimbangan jumlah sampel antara kelas *benign* dan *malignant* pada beberapa tingkat pembesaran.

Setelah *preprocessing*, data dibagi menjadi subset *training* dan *testing* dengan proporsi 80:20 pada setiap tingkat pembesaran (40×, 100×, 200×, dan 400×). Masing-masing subset kemudian digunakan untuk melatih tiga model CNN, yaitu *VGG19*, *ResNet50*, dan *EfficientNetB0*. Ketiga model menerapkan pendekatan *transfer learning* berbasis *feature extraction*, di mana *backbone* pretrained *ImageNet* digunakan untuk mengekstraksi fitur visual penting dari jaringan histopatologi, sementara lapisan klasifikasi baru dilatih selama 15 epoch menggunakan optimizer Adam. Pelatihan dilakukan secara terpisah pada keempat tingkat pembesaran sehingga karakteristik visual unik di setiap skala dapat dipelajari secara mandiri oleh masing-masing model.

Tahap berikutnya adalah proses ensemble *soft voting*. Pada tahap ini, setiap model menghasilkan nilai probabilitas untuk kedua kelas (*benign* dan *malignant*). Probabilitas tersebut kemudian dirata-ratakan sehingga keputusan akhir diambil berdasarkan konsensus tingkat kepercayaan antar-model. Pendekatan ensemble ini dirancang untuk memanfaatkan kekuatan komplementer ketiga arsitektur: kemampuan *VGG19* dalam menangkap tekstur halus, stabilitas gradien dari *ResNet50*, serta efisiensi representasi multi-skala dari *EfficientNetB0*. Dengan demikian, ensemble berfungsi sebagai mekanisme penyatuan prediksi untuk meningkatkan stabilitas dan mengurangi bias model individu terhadap variasi pembesaran maupun tekstur jaringan.

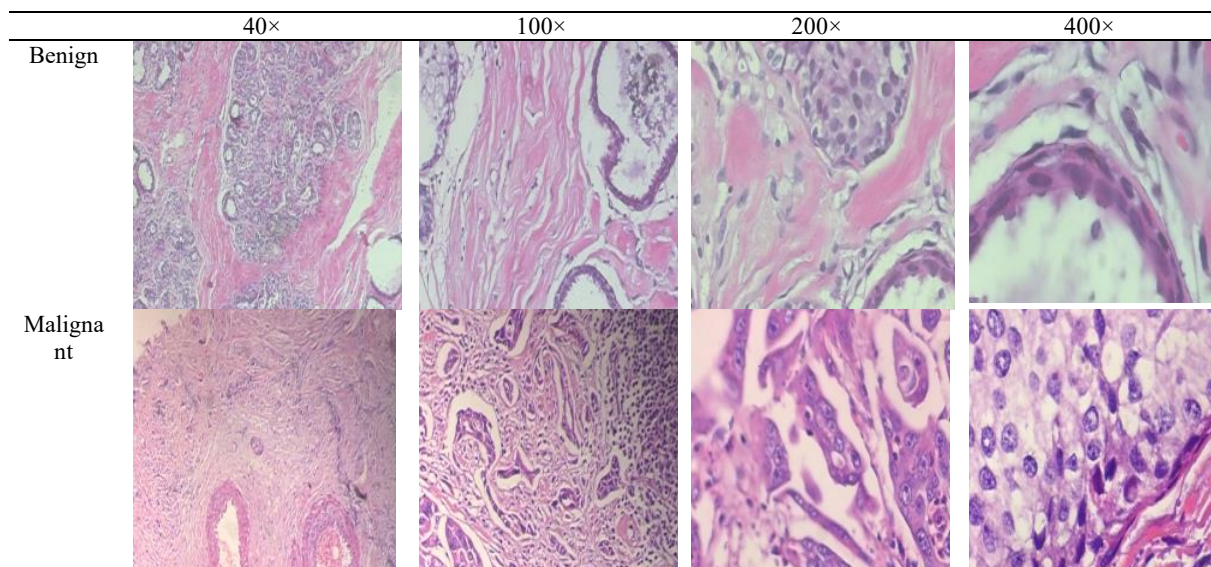
Tahap terakhir adalah evaluasi kinerja menggunakan metrik akurasi, precision, recall, dan F1-score, serta analisis melalui *confusion matrix* untuk mengamati pola kesalahan klasifikasi. Evaluasi dilakukan pada empat tingkat pembesaran untuk menilai kemampuan generalisasi model di berbagai skala citra. Hasil-hasil tersebut kemudian dianalisis dan dibandingkan dengan performa model tunggal untuk menentukan efektivitas pendekatan ensemble dalam menyelesaikan masalah klasifikasi citra histopatologi kanker payudara.

#### 3.1 Proses Pengumpulan dan Pemahaman Data

Dataset yang menjadi fondasi dalam penelitian ini adalah BreaKHis (Breast Cancer Histopathological Image Classification), sebuah dataset standar industri yang memuat citra histopatologi jaringan payudara hasil biopsi bedah. Dataset ini terbagi secara dikotomis ke dalam dua kategori klinis utama, yaitu *benign* (tumor jinak) dan *malignant* (tumor ganas/kanker). Secara hierarkis, kedua kategori tersebut mencakup delapan jenis sub-tipe tumor yang berbeda, sehingga memberikan spektrum data yang kaya untuk pengujian model klasifikasi. Salah satu karakteristik unik dari dataset BreaKHis adalah penyediaan citra pada empat tingkat perbesaran mikroskopis yang berbeda: 40×, 100×, 200×, dan 400×. Variasi perbesaran ini menghasilkan diversitas visual yang signifikan, di mana tingkat perbesaran rendah (40× dan 100×) lebih menonjolkan arsitektur jaringan secara makroskopis, sementara perbesaran tinggi (200× dan 400×) mengekspos detail morfologi seluler dan pola inti sel secara mikroskopis. Keberagaman skala ini memungkinkan arsitektur deep learning yang dikembangkan untuk mempelajari fitur-fitur diskriminatif pada berbagai tingkat kedalaman (multiskala), yang sangat krusial karena tanda-tanda keganasan sering kali muncul pada detail tekstur yang halus. Namun, perbedaan detail visual yang kontras antar tingkat pembesaran ini juga menghadirkan tantangan teknis tersendiri terkait dengan invariansi skala dan generalisasi model. Representasi visual dari

perbandingan karakteristik citra untuk setiap tingkat pembesaran pada kelas benign dan malignant dipaparkan secara mendetail pada Tabel 3 di bawah ini.

**Tabel 3.** Data yang digunakan



Proses pengumpulan data dilakukan dengan mengunduh dataset dari platform Kaggle dalam struktur folder per-zoom. Setiap citra berada dalam format RGB dengan ukuran bervariasi, sehingga diperlukan proses standarisasi pada tahap pra-pengolahan. Sebelum model CNN dilatih, seluruh citra melalui beberapa langkah *preprocessing* agar konsisten secara dimensi dan kualitas. Tahapan tersebut meliputi:

- Resize* citra ke 224×224 piksel untuk menyesuaikan kebutuhan arsitektur CNN modern.
- Normalisasi piksel ke rentang [0, 1] untuk mempercepat proses konvergensi model.
- Augmentasi* data seperti rotation, horizontal/vertical flip, zoom-range, dan brightness adjustment untuk memperkaya variasi data serta mengurangi risiko *overfitting*.
- Pembagian data menjadi *train-validation-test* per-zoom agar proses evaluasi tetap seimbang dan tidak menyebabkan data *leakage*.

Selain itu, dataset BreakHis memiliki ketidakseimbangan internal pada beberapa tingkat pembesaran, terutama pada kelas *benign* yang jumlahnya lebih sedikit. Oleh karena itu, *augmentasi* diterapkan secara adaptif untuk menyeimbangkan distribusi di tingkat zoom tertentu. Proses persiapan ini memastikan bahwa seluruh model CNN yang digunakan pada tahap berikutnya menerima input yang seragam sehingga performa dapat dibandingkan secara adil, sekaligus memberikan landasan yang kuat untuk tahap ensemble yang menjadi fokus utama penelitian.

### 3.2 Proses Pengolahan Data dan Hasil Model CNN

Pengolahan data pada penelitian ini dimulai setelah seluruh citra dari empat tingkat pembesaran (40×, 100×, 200×, 400×) dipisahkan ke dalam kelas *benign* dan *malignant*. Tahap *preprocessing* yang telah dilakukan mencakup *resize* citra ke 224×224 piksel, normalisasi intensitas piksel, serta augmentasi menggunakan *horizontal flip*, *rotation*, *zooming*, dan *brightness shift*. Tahap ini bertujuan untuk meningkatkan keragaman data pelatihan dan membantu model mengatasi variasi warna dan struktur jaringan yang lazim muncul pada citra *histopathology*. Selain itu, proses *preprocessing* berperan penting dalam menstabilkan distribusi data sehingga model tidak terlalu sensitif terhadap perbedaan pencahayaan dan pewarnaan jaringan. Standarisasi ukuran citra juga memastikan kompatibilitas input dengan seluruh arsitektur CNN yang digunakan. Dengan demikian, fitur yang diekstraksi pada tahap awal lebih merepresentasikan karakteristik biologis jaringan dibandingkan artefak visual.

Setelah tahap *preprocessing* selesai, data yang telah distandarasi kemudian dipisahkan berdasarkan tingkat pembesaran untuk memastikan setiap skala dianalisis secara independen. Pendekatan ini dilakukan karena setiap tingkat pembesaran memiliki karakteristik visual dan kompleksitas fitur yang berbeda. Dengan memisahkan data per-zoom, proses pelatihan model dapat lebih terfokus dalam mempelajari pola spesifik pada masing-masing skala citra. Selain itu, strategi ini memungkinkan evaluasi performa model secara lebih objektif terhadap variasi struktur jaringan *histopathology*. Pendekatan per-skala ini juga memudahkan analisis komparatif antar pembesaran untuk mengetahui sensitivitas model terhadap detail mikroskopis maupun pola jaringan global. Tahap ini menjadi penghubung penting antara proses pra-pengolahan data dan tahap pelatihan model CNN yang dilakukan selanjutnya.

Setiap subset pembesaran kemudian diproses menggunakan tiga arsitektur CNN berbeda, yaitu *EfficientNetB0*, *ResNet50*, dan *VGG19*, yang dilatih secara terpisah menggunakan konfigurasi *learning rate*, *batch size*, dan *optimizer Adam* yang seragam agar performa tiap model dapat dibandingkan secara adil. Proses pelatihan dilakukan sebanyak 15 *epoch* pada seluruh model dan seluruh tingkat pembesaran. Hasil pelatihan menunjukkan bahwa ketiga model mampu membangun pola representasi fitur yang cukup kuat meskipun memiliki karakteristik arsitektur yang berbeda.

Perbedaan hasil yang diperoleh mencerminkan pengaruh desain arsitektur terhadap kemampuan model dalam menangkap variasi tekstur dan struktur sel kanker. Kinerja masing-masing model CNN pada seluruh tingkat pembesaran ditunjukkan pada Tabel 4, yang memuat metrik akurasi, presisi, *recall*, dan *F1-score*.

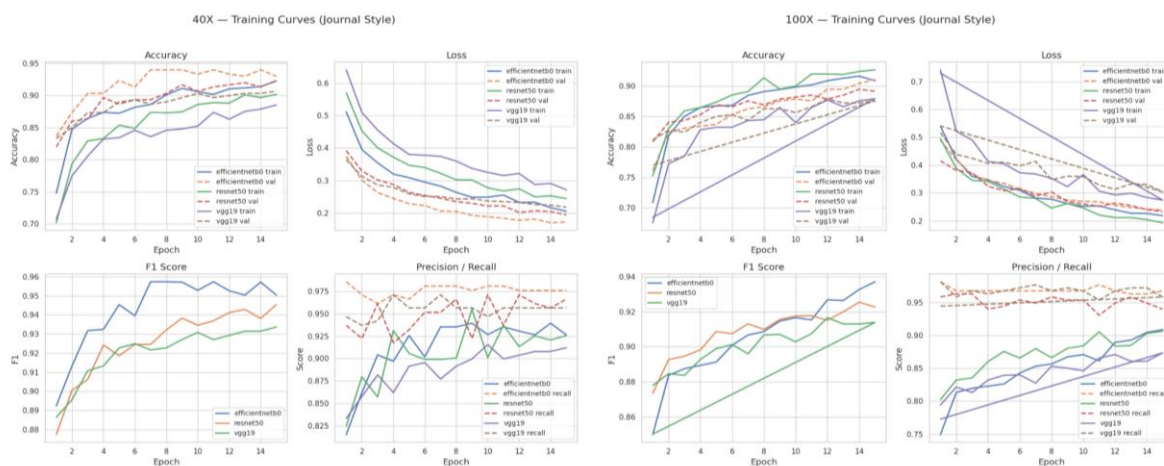
**Tabel 4.** Hasil Pengolahan

Zoom	Model	Accuracy	Precision	Recall	F1 - Score
40×	EfficientNetB0	0.922581	0.926267	0.975728	0.950355
	ResNet50	0.901792	0.925581	0.966019	0.945368
	VGG19	0.885305	0.912037	0.956311	0.933649
100×	EfficientNetB0	0.916151	0.904348	0.962963	0.932735
	ResNet50	0.926461	0.906250	0.939815	0.922727
	VGG19	0.879038	0.873418	0.958333	0.913907
200×	EfficientNetB0	0.904897	0.915888	0.942308	0.928910
	ResNet50	0.931867	0.951923	0.951923	0.951923
	VGG19	0.877218	0.869565	0.961538	0.913242
400×	EfficientNetB0	0.892380	0.928177	0.908108	0.918033
	ResNet50	0.910448	0.928177	0.908108	0.918033
	VGG19	0.844462	0.854369	0.951351	0.900256

Secara umum, ketiga model menunjukkan performa yang stabil pada seluruh tingkat pembesaran, dengan *EfficientNetB0* dan *ResNet50* menjadi model yang paling konsisten memiliki skor tertinggi. *ResNet50* mencapai performa terbaik pada sebagian besar pembesaran berkat arsitektur residual yang membantu proses *gradient flow* lebih stabil, sementara *EfficientNetB0* unggul pada data bertekstur halus seperti pembesaran 400× karena pendekatan *compound scaling* yang efisien dalam menangkap detail mikro. *VGG19*, meskipun tertinggal dari dua model lainnya, tetap menunjukkan performa kompetitif untuk klasifikasi dua kelas, namun cenderung kurang stabil pada validation loss terutama di pembesaran tinggi. Hal ini terkait dengan kedalaman arsitektur tanpa mekanisme *skip connection* sehingga lebih sensitif terhadap *overfitting*.

Perbedaan performa antar model ini menunjukkan bahwa karakteristik arsitektur CNN sangat berpengaruh terhadap kemampuan generalisasi pada citra histopatologi. Model dengan mekanisme residual dan scaling adaptif cenderung lebih robust dalam menangani variasi tekstur dan kompleksitas struktur jaringan. Selain itu, stabilitas nilai recall dan F1-score pada *EfficientNetB0* dan *ResNet50* mengindikasikan kemampuan deteksi kelas malignant yang lebih konsisten. Temuan ini penting mengingat kesalahan klasifikasi pada kelas malignant dapat berdampak signifikan dalam konteks klinis. Oleh karena itu, analisis performa tidak hanya berfokus pada akurasi, tetapi juga pada keseimbangan metrik evaluasi secara keseluruhan.

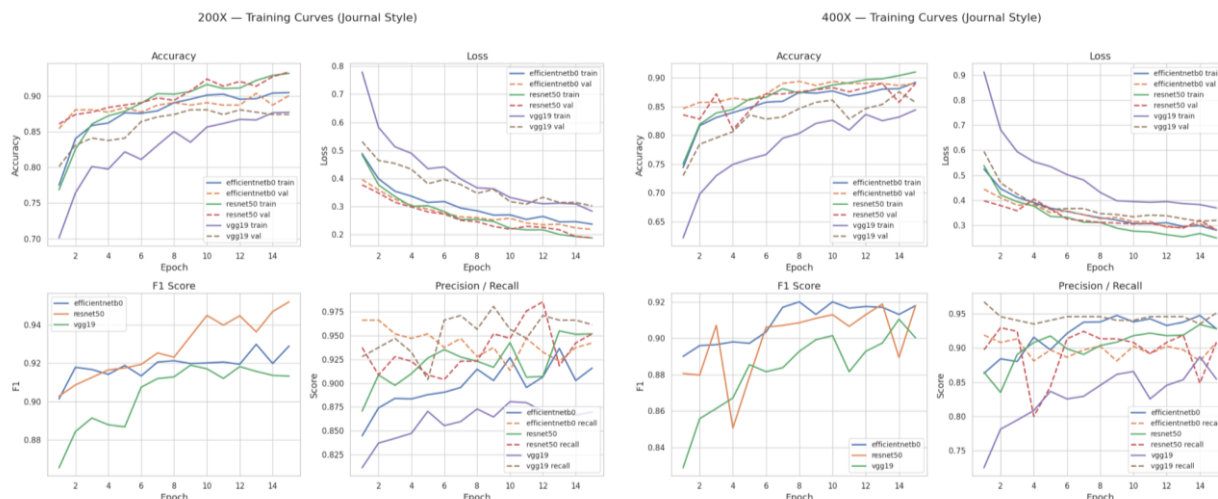
Untuk memahami dinamika pelatihan, seluruh metrik divisualisasikan dalam bentuk grafik yang mencakup akurasi, loss, F1-score, precision, dan recall sepanjang 15 *epoch*. Visualisasi ini digunakan untuk mengamati proses konvergensi model serta kestabilan pembelajaran selama pelatihan berlangsung. Melalui analisis kurva metrik, dapat dievaluasi hubungan antara peningkatan akurasi dan penurunan loss sebagai indikator efektivitas optimisasi model. Selain itu, pola perubahan precision, recall, dan F1-score memberikan gambaran keseimbangan performa model dalam mendeteksi kelas benign dan malignant secara konsisten pada setiap *epoch*. Pola perubahan metrik untuk tingkat pembesaran rendah (40× dan 100×) ditampilkan pada Gambar 2 berikut.



**Gambar 2.** Kurva accuracy, loss, precision, recall, dan F1-score untuk pembesaran 40× dan 100×.

Berdasarkan Gambar 2, seluruh model menunjukkan peningkatan akurasi yang konsisten sejak epoch awal. *ResNet50* dan *EfficientNetB0* memperlihatkan konvergensi yang lebih stabil dibandingkan *VGG19*, yang mengalami fluktuasi terutama pada *validation loss*. Selain itu, F1-score pada kedua pembesaran tersebut menunjukkan bahwa

*EfficientNetB0* memiliki keseimbangan precision–recall yang paling stabil. Untuk tingkat pembesaran lebih tinggi (200× dan 400×), pola dinamika pelatihan disajikan pada Gambar 3.



**Gambar 3.** Kurva accuracy, loss, precision, recall, dan F1-score untuk pembesaran 200× dan 400×.

Mengacu pada Gambar 3, performa pelatihan pada pembesaran tinggi masih mempertahankan pola yang serupa dengan pembesaran rendah. *ResNet50* kembali menunjukkan stabilitas terbaik, dengan kurva *training* dan *validation* yang relatif halus dan konsisten. Pada pembesaran 400×, *EfficientNetB0* mencapai F1-score tertinggi karena arsitekturnya lebih adaptif terhadap detail tekstur halus yang muncul pada zoom tinggi. Sebaliknya, *VGG19* menampilkan variabilitas yang lebih besar karena tidak memiliki residual block, sehingga proses pembelajaran menjadi lebih sensitif terhadap perubahan gradien dan lebih mudah mengalami fluktuasi. Secara keseluruhan, peningkatan akurasi di semua model cenderung stabil dan mencapai konvergensi sekitar epoch ke-12. *Validation loss* yang lebih rendah daripada *training loss* pada *EfficientNetB0* dan *ResNet50* menunjukkan kemampuan generalisasi yang baik. Precision dan recall juga relatif seimbang, menandakan bahwa model tidak condong memihak salah satu kelas (*benign* atau *malignant*) dan mampu memberikan prediksi yang konsisten di berbagai tingkat pembesaran. Selain itu, pola kurva pada pembesaran tinggi juga menunjukkan bahwa kompleksitas fitur mikroskopis tidak sepenuhnya mengganggu proses pembelajaran model. Hal ini terlihat dari konsistensi nilai F1-score dan stabilitas *validation accuracy* pada *ResNet50* dan *EfficientNetB0*. Kinerja yang tetap terjaga ini mengindikasikan bahwa kedua model memiliki kapasitas representasi yang cukup untuk menangkap variasi struktur sel kanker, seperti pola nukleus, distribusi kromatin, dan bentuk jaringan. Di sisi lain, meskipun *VGG19* masih mampu mencapai akurasi yang kompetitif, ketidastabilannya menunjukkan bahwa arsitektur konvensional yang dalam tanpa residual connection mulai mengalami keterbatasan pada tingkat pembesaran tinggi. Fluktuasi yang muncul pada *training curve* menandakan bahwa model lebih mudah terpengaruh oleh noise atau variasi fitur lokal yang sangat kecil, sehingga hasilnya kurang konsisten dibandingkan dua model lainnya yang memiliki mekanisme stabilisasi internal.

### 3.3 Proses Ensemble Multi-Arsitektur CNN

Proses ensemble pada penelitian ini dilakukan dengan menggabungkan keluaran dari tiga model CNN yang telah dilatih secara terpisah, yaitu *EfficientNetB0*, *ResNet50*, dan *VGG19*. Setiap model menghasilkan skor probabilitas untuk masing-masing kelas, kemudian seluruh probabilitas ini digabungkan menggunakan metode *soft voting*. Pada pendekatan ini, probabilitas dari ketiga model dirata-ratakan sehingga prediksi akhir diperoleh berdasarkan konsensus tingkat kepercayaan, bukan sekadar jumlah suara.

Sebelum dilakukan penggabungan, seluruh model diseragamkan pada tahap pra-pemrosesan input dan normalisasi keluaran agar kontribusi tiap arsitektur tetap seimbang. Setiap model tetap mempertahankan karakteristik arsitekturnya seperti residual connection pada *ResNet50* atau struktur depth yang dalam pada *VGG19* namun keputusan akhirnya tetap diambil secara kolektif melalui mekanisme ensemble.

Dengan konfigurasi ini, ensemble berfungsi sebagai tahap penguatan prediksi, di mana variasi antar-model dimanfaatkan untuk mengurangi ketergantungan pada satu arsitektur saja. Tahapan ini menjadi langkah penting sebelum dilakukan evaluasi kinerja ensemble pada bagian selanjutnya.

### 3.4 Evaluasi Ensemble

Evaluasi akhir dilakukan dengan menggabungkan prediksi tiga model dasar *EfficientNetB0*, *ResNet50*, dan *VGG19* menggunakan pendekatan *soft voting*. Hasil ensemble kemudian dibandingkan langsung dengan performa masing-masing model pada empat tingkat pembesaran citra, yaitu 40×, 100×, 200×, dan 400×. Pendekatan ensemble terbukti mampu menghasilkan akurasi yang lebih stabil dan konsisten dibanding model tunggal, terutama karena proses agregasi probabilitas dapat mereduksi bias spesifik arsitektur serta menyeimbangkan kekuatan antar model.

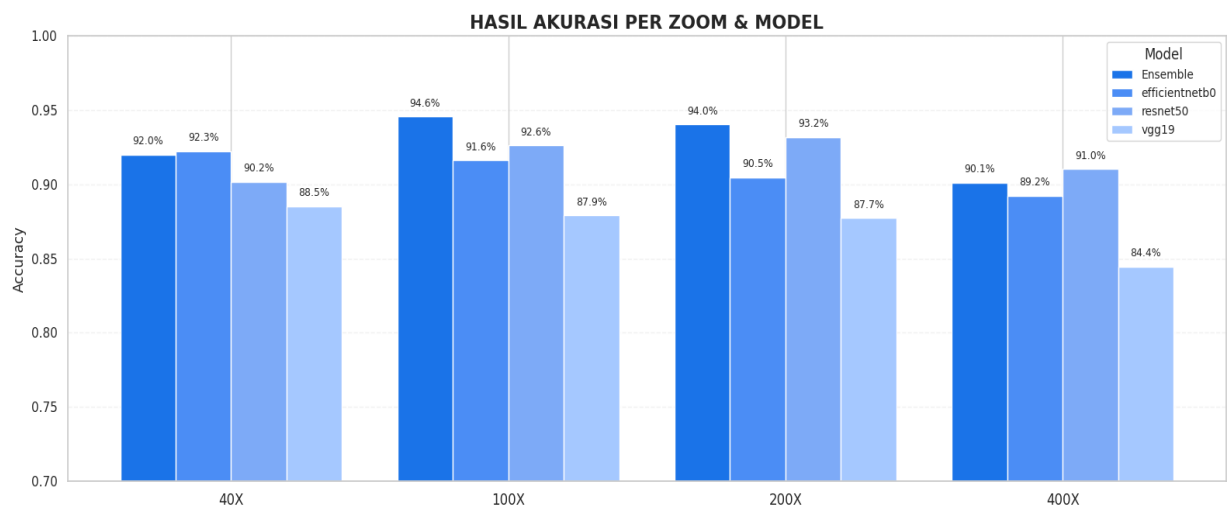
Hasil lengkap performa ensemble ditampilkan pada Tabel berikut, yang menunjukkan nilai akurasi, precision, recall, serta F1-score untuk tiap tingkat zoom. Secara umum, metode ensemble memberikan performa terbaik pada zoom 100× dengan akurasi 94.56% dan F1-score 96,17%, sementara performa terendah berada pada zoom 400× dengan akurasi tetap berada pada tingkat kompetitif yaitu 90.10%. Pola ini menunjukkan bahwa ensemble sangat efektif pada tingkat pembesaran yang memiliki detail jaringan lebih seimbang antara struktur makro dan mikro. Performa lengkap metode ensemble pada keempat tingkat pembesaran ditunjukkan pada Tabel 5, meliputi akurasi, presisi, recall, dan F1-score.

**Tabel 5.** Hasil Evaluasi Ensemble

Zoom	Model	Accuracy	Precision	Recall	F1-Score
40×	Ensemble	0.920000	0.902655	0.990291	0.944444
100×	Ensemble	0.945687	0.934498	0.990741	0.961798
200×	Ensemble	0.940397	0.940092	0.976077	0.957746
400×	Ensemble	0.901099	0.907216	0.951351	0.928760

Berdasarkan hasil pada Tabel 5, terlihat bahwa metode ensemble secara konsisten mampu mempertahankan keseimbangan antara precision dan recall pada seluruh tingkat pembesaran. Nilai recall yang tinggi pada setiap zoom menunjukkan bahwa pendekatan ensemble efektif dalam meminimalkan kesalahan false negative, yang sangat krusial dalam deteksi kanker payudara. Peningkatan nilai F1-score dibandingkan sebagian besar model tunggal mengindikasikan bahwa proses penggabungan probabilitas berhasil menyeimbangkan sensitivitas dan ketepatan prediksi. Stabilitas performa ini menunjukkan bahwa ensemble mampu mereduksi fluktuasi prediksi yang biasanya muncul akibat perbedaan karakteristik dan bias masing-masing arsitektur CNN. Selain itu, performa yang relatif konsisten pada keempat tingkat pembesaran menunjukkan bahwa ensemble memiliki kemampuan generalisasi lintas skala yang lebih baik dibandingkan model individual. Dengan demikian, penerapan ensemble *soft voting* memberikan kontribusi signifikan dalam meningkatkan keandalan dan konsistensi sistem klasifikasi citra histopatologi kanker payudara.

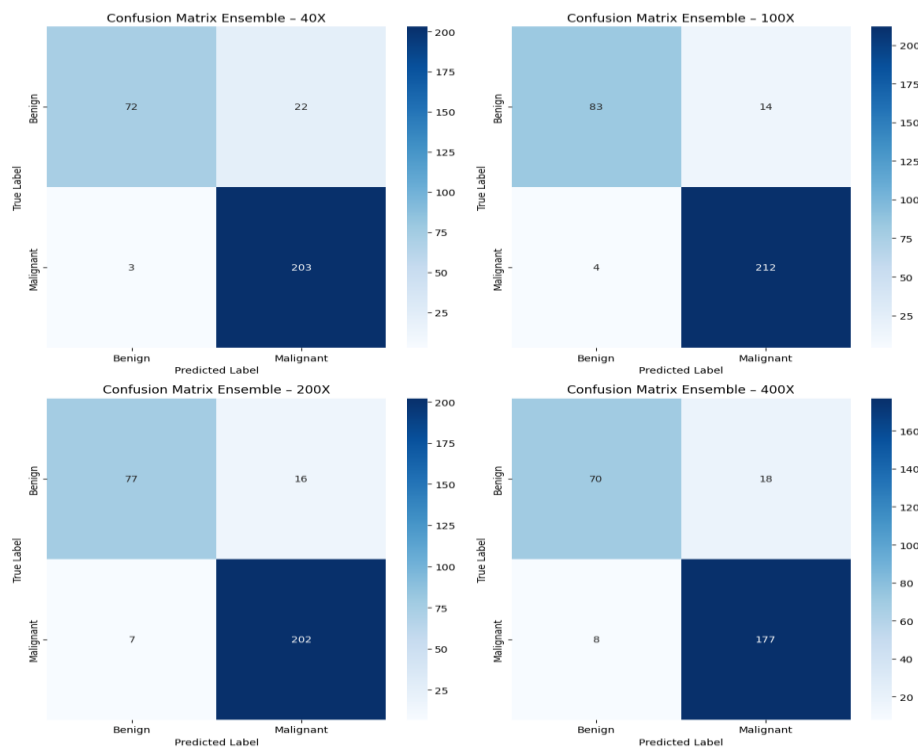
Grafik berikut menampilkan perbandingan akurasi antara tiga model CNN serta hasil ensemble pada seluruh tingkat pembesaran. Terlihat bahwa ensemble selalu menghasilkan nilai yang berada di posisi teratas atau setidaknya sangat kompetitif, dengan margin peningkatan yang paling jelas pada pembesaran 100× dan 200×. Perbandingan akurasi model tunggal dan ensemble pada seluruh tingkat pembesaran divisualisasikan pada Gambar 4 berikut.



**Gambar 4.** Diagram Hasil Akurasi Model

Dari diagram di atas dapat dilihat bahwa ensemble consistently berada pada tingkat akurasi tertinggi di hampir seluruh pembesaran. Pada zoom 40×, ensemble mengimbangi performa EfficientNetB0 dengan akurasi sekitar 92%. Pada zoom 100× dan 200×, ensemble melampaui seluruh model tunggal, menunjukkan bahwa kombinasi tiga arsitektur mampu menangkap pola tekstur histopatologi secara lebih stabil. Sementara pada zoom 400× jenis citra yang memiliki variasi tekstur paling tinggi ensemble tetap mempertahankan akurasi di atas 90%, lebih tinggi dibanding VGG19 dan sedikit lebih baik dari ResNet50.

Selain evaluasi berbasis metrik akurasi, precision, recall, dan F1-score, performa model ensemble juga dianalisis menggunakan *confusion matrix* pada setiap tingkat pembesaran 40×, 100×, 200×, dan 400×. *Confusion matrix* memberikan gambaran lebih rinci mengenai jenis kesalahan yang dilakukan model, terutama terkait false positive dan false negative, yang sangat penting dalam konteks klasifikasi citra histopatologi. Visualisasi ini membantu memastikan bahwa peningkatan metrik bukan hanya terjadi secara agregat, tetapi juga konsisten pada distribusi prediksi per kelas. Detail kesalahan klasifikasi pada tiap tingkat pembesaran divisualisasikan pada Gambar 5 dalam bentuk *confusion matrix*.



Gambar 5. Confusion Matrix Ensemble

Dari keempat *confusion matrix* tersebut terlihat bahwa tingkat kesalahan false negative berada pada kisaran yang sangat rendah (3–8 kasus), yang menunjukkan bahwa model ensemble sangat sensitif dalam mendeteksi citra malignant. Sementara itu, false positive sedikit lebih tinggi pada pembesaran rendah seperti 40× dan pembesaran tinggi seperti 400×, mengindikasikan bahwa variasi struktur jaringan pada dua tingkat ini dapat menyebabkan model lebih sulit membedakan pola *benign* dan *malignant*. Meskipun demikian, tingkat prediksi benar tetap mendominasi, sehingga mendukung kesimpulan bahwa ensemble memberikan stabilitas yang lebih baik dibanding model tunggal. Kinerja precision dan recall yang seimbang menunjukkan bahwa ensemble tidak bias terhadap kelas *benign* maupun *malignant*, sementara F1-score yang tinggi terutama pada zoom 100× dan 200× mengindikasikan bahwa ensemble sangat efektif dalam situasi di mana detail jaringan cukup kaya untuk mendukung keputusan klasifikasi gabungan antar model.

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan ensemble multi-arsitektur CNN mampu memberikan performa klasifikasi citra histopatologi kanker payudara yang lebih stabil dan akurat dibandingkan model tunggal. *EfficientNetB0* tampil paling unggul pada zoom 40× dengan akurasi 0.922581 dan F1-score 0.950535, sementara *ResNet50* menunjukkan performa terbaik pada zoom 200× dengan akurasi 0.931867 dan F1-score 0.951923, mengindikasikan bahwa tiap arsitektur memiliki kemampuan ekstraksi fitur yang optimal pada karakteristik skala tertentu. Ketika ketiga model digabungkan melalui *soft voting* ensemble, peningkatan performa terlihat konsisten pada seluruh tingkat pembesaran, dengan hasil terbaik pada zoom 100× berupa akurasi 0.945687 dan F1-score 0.961798. Ensemble juga menunjukkan sensitivitas tinggi terhadap kelas malignant melalui nilai recall yang sangat dominan, sekaligus menjaga keseimbangan precision sehingga prediksi tetap stabil meskipun variasi tekstur jaringan sangat beragam. Temuan ini menegaskan bahwa penggabungan beberapa arsitektur mampu memperkuat representasi fitur dan mengatasi kelemahan model tunggal. Penelitian berikutnya berpotensi diarahkan pada pengembangan arsitektur lebih modern seperti *ConvNeXt*, *EfficientNetV2*, atau *Vision Transformer*, penerapan teknik ensemble lanjutan seperti stacking untuk memperoleh prediksi yang lebih adaptif, serta eksplorasi klasifikasi multi-kelas atau pemanfaatan model generatif untuk memperkaya variasi data sehingga sistem diagnosis berbasis citra medis dapat mencapai tingkat presisi yang lebih tinggi.

#### REFERENCES

- [1] I. Haq *et al.*, "A review of breast cancer histopathology image analysis with deep learning: Challenges, innovations, and clinical integration," *Image and Vision Computing*, vol. 162, Art. no. 105708, 2025, doi: 10.1016/j.imavis.2024.105708.
- [2] M. A. Talukder, "An improved XAI-based DenseNet model for breast cancer detection using reconstruction and fine-tuning," *Results in Engineering*, vol. 26, Art. no. 104802, 2025, doi: 10.1016/j.rineng.2024.104802.



- [3] M. K. Nath, K. Sundararajan, S. Mathivanan, dan B. Thandapani, "Analysis of breast cancer classification and segmentation techniques: A comprehensive review," *Informatics in Medicine Unlocked*, vol. 56, Art. no. 101642, 2025, doi: 10.1016/j.imu.2025.101642.
- [4] K. Gupta dan N. Chawla, "Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN," *Procedia Computer Science*, vol. 167, pp. 878–889, 2020, doi: 10.1016/j.procs.2020.03.427.
- [5] Y. Zheng *et al.*, "Application of transfer learning and ensemble learning in image-level classification for breast histopathology," *Intelligent Medicine*, vol. 3, no. 2, pp. 115–128, 2023, doi: 10.1016/j.imed.2022.12.001.
- [6] M. K. B. Lawal *et al.*, "Artificial intelligent-powered detection of breast cancer," *Journal of Radiation Research and Applied Sciences*, vol. 18, no. 1, Art. no. 101422, 2025, doi: 10.1016/j.jrras.2024.101422.
- [7] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. J. Aguirre, dan A. M. Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," *Sensors*, vol. 20, no. 16, Art. no. 4373, 2020, doi: 10.3390/s20164373.
- [8] R. B. Eshun, M. Bikdash, dan A. K. M. K. Islam, "A deep convolutional neural network for the classification of imbalanced breast cancer dataset," *Healthcare Analytics*, vol. 5, Art. no. 100330, 2024, doi: 10.1016/j.health.2024.100330.
- [9] A. H. M. Z. Karim, K. B. O. Mahmud, dan C. Shahnaz, "A hybrid feature extraction based ensemble model for breast cancer detection and classification using different medical images," *Intelligence-Based Medicine*, vol. 12, Art. no. 100290, 2025, doi: 10.1016/j.ibmed.2024.100290.
- [10] M. Fatima, R. Zia, dan I. A. Usmani, "A hybrid framework integrating deep learning and XAI for breast cancer detection with stratified cross-validation," *Soft Computing Letters*, Art. no. 100231 (in press), 2025, doi: 10.1016/j.sasc.2025.200412.
- [11] M. M. Rahaman, E. K. A. Millar, dan E. Meijering, "Generalized deep learning for histopathology image classification using supervised contrastive learning," *Journal of Advanced Research*, vol. 75, pp. 389–404, 2025, doi: 10.1016/j.jare.2024.06.014.
- [12] S. Singh *et al.*, "Hybrid models for breast cancer detection via transfer learning technique," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 2655–2672, 2023, doi: 10.32604/cmc.2023.032363.
- [13] D. Bhulakshmi dan D. S. Rajput, "Leveraging ensembles of pre-trained CNNs for improved lung cancer detection and classification," *Procedia Computer Science*, vol. 260, pp. 151–159, 2025, doi: 10.1016/j.procs.2024.11.026.
- [14] G. Wang *et al.*, "Multi-classification of breast cancer pathology images based on a two-stage hybrid network," *Journal of Cancer Research and Clinical Oncology*, vol. 150, no. 11, Art. no. 505, 2024, doi: 10.1007/s00432-024-06002-y.
- [15] S. I. Khan, A. Shahrior, R. Karim, M. Hasan, dan A. Rahman, "MultiNet: A deep neural network approach for detecting breast cancer through multi-scale feature fusion," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 6217–6228, 2022, doi: 10.1016/j.jksuci.2021.08.004.
- [16] M. M. Rahman, M. S. I. Khan, dan H. M. H. Babu, "BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer," *Array*, vol. 16, Art. no. 100256, 2022, doi: 10.1016/j.array.2022.100256.
- [17] S. S. M. Khairi *et al.*, "Deep learning on histopathology images for breast cancer classification: A bibliometric analysis," *Healthcare*, vol. 10, no. 1, Art. no. 10, 2022, doi: 10.3390/healthcare10010010.
- [18] A. Rafiq *et al.*, "Detection and classification of histopathological breast images using a fusion of CNN frameworks," *Diagnostics*, vol. 13, no. 10, Art. no. 1700, 2023, doi: 10.3390/diagnostics13101700.
- [19] A. A. Balasubramanian *et al.*, "Ensemble deep learning-based image classification for breast cancer subtype and invasiveness diagnosis from whole slide image histopathology," *Cancers*, vol. 16, no. 12, Art. no. 2222, 2024, doi: 10.3390/cancers16122222.
- [20] Y. Çetin-Kaya, "Equilibrium optimization-based ensemble CNN framework for breast cancer multiclass classification using histopathological image," *Diagnostics*, vol. 14, no. 19, Art. no. 2253, 2024, doi: 10.3390/diagnostics14192253.
- [21] Z. Gan dan A. Subasi, "Promoting accuracy in low-magnification histopathology grading: With augmentation and multi-dilation model," *Biomedical Signal Processing and Control*, vol. 86, Art. no. 105118, 2023, doi: 10.1016/j.bspc.2023.105118.