

Klasifikasi Citra Histologi Kanker Payudara Menggunakan Metode Ensemble CNN

Galang Aji Mahesa^{*1}, Agus Eko Minarno², Yufis Azhar³

^{1, 2, 3}Universitas Muhammadiyah Malang

galangamahesa@gmail.com^{*1}, aguseko@umm.ac.id², yufis@umm.ac.id³

Abstrak

Kanker payudara merupakan jenis penyakit serius dan menjadi penyebab kematian paling tinggi pada wanita. Kanker payudara dapat didiagnosis melalui citra tes medis, seperti citra radiologi dan hispatologi. Namun, karena citra histologi memiliki kompleksitas dan keberagaman, pemeriksaan secara manual memiliki kelemahan, yaitu, membutuhkan keahlian yang tinggi dan memakan waktu serta rawan kesalahan. Deep learning telah banyak diterapkan pada klasifikasi histologi termasuk kanker payudara, dengan otomatisasi dapat membantu mengatasi kekurangan pada metode diagnosis manual. Pada penelitian ini, dikembangkan Deep learning menggunakan CNN dengan metode ensemble. Model MobileNet, MobileNetV2, dan VGG16 dilatih terhadap dataset dan dilakukan averaging terhadap hasil kombinasi model model tersebut. Hasil percobaan menunjukkan peningkatan pada ensemble dibandingkan tiap model dengan balanced accuracy 0.8689 dan F1-score 0.8682 pada kombinasi VGG16 + MobileNet, hasil tersebut memberikan peningkatan dibandingkan penelitian sebelumnya.

Kata kunci: CNN, Breast Cancer, Ensemble, MobileNet, MobileNetV2

Abstract

Breast cancer is a serious disease and the leading cause of death in women. Breast cancer can be diagnosed through medical imaging tests, such as radiological and hispathological images. However, because histological images have complexity and diversity, the manual examination has disadvantages, namely, requiring high expertise, time consuming, and prone to errors. Deep learning has been widely applied to histological classification including breast cancer, with automation to help overcome the shortcomings of manual diagnostic methods. In this study, Deep learning was developed using CNN with the ensemble method. The MobileNet, MobileNetV2 and VGG16 models were trained on the dataset and averaged the results of the two models. The experimental results show an increase in the ensemble compared to each model with an balanced accuracy 0.8689 dan F1-score 0.8682 in VGG16 + MobileNet combination, these results provide an increase compared to previous studies.

Keywords: CNN, Breast Cancer, Ensemble, MobileNet, MobileNetV2

1. Pendahuluan

Kanker payudara adalah urutan kanker teratas yang mempengaruhi wanita serta kanker penyebab kematian paling tinggi pada wanita (1). Kanker payudara merupakan penyakit yang serius, dapat dimulai dari hampir semua jaringan tubuh dan menyerang bagian tubuh yang berdekatan atau organ lain. Berdasarkan data yang didapatkan dari press release No 292 International Agency for Research on Cancer di 185 negara dengan 36 jenis kanker, beban kanker dunia diperkirakan meningkat ke 19.3 juta kasus baru dan 10 juta kematian pada 2020. Dengan kanker payudara yang paling banyak muncul dengan 11.7% dari total kasus baru, dan 6.9% dari semua kanker penyebab kematian (2). Kanker payudara bisa didiagnosis dengan citra tes medis, seperti citra radiologi dan hispatologi. Dengan citra hispatologi, ahli hispatologi akan memeriksa secara visual keteraturan bentuk sel dan distribusi jaringan serta menentukan daerah kanker dan keganasannya (3). Namun, karena citra histologi memiliki kompleksitas dan keberagaman, pemeriksaan secara manual tentunya akan membutuhkan keahlian yang tinggi dan memakan waktu serta rawan kesalahan (4).

Jumlah ahli dibidang ini juga sedikit, sebagai contoh di daerah sub-sahara Afrika terdapat satu tenaga ahli untuk satu juta orang, dan di china satu tenaga ahli untuk 134.571 orang (5). Dengan perkembangan pemrosesan citra dan pembelajaran mesin, mengembangkan system

penegenalan pola kanker payudara menjadi kebutuhan untuk membantu para ahli mengoptimalkan kualitas diagnostic (3).

Beberapa tahun belakangan ini, klasifikasi citra semakin berkembang dan menjadi salah satu topik yang sering dibicarakan dibidang teknologi. Misalnya dibidang e-commerce, kesehatan, media sosial, dan sebagainya, hal ini juga didukung oleh berkembangnya data dari berbagai bidang (6). Klasifikasi citra, merupakan salah satu contoh penerapan dari machine learning yang mendasar dan paling banyak digunakan dibidang computer vision. Beberapa tahun terakhir, convolution neural network (CNN) menunjukkan keberhasilan dalam penerapan klasifikasi citra karena kecepatan dan fungsi ekstraksi fitur yang akurat serta end-to-end trainable network framework (7).

Araújo, et. al., pada tahun 2017 juga melakukan penelitian terkait kanker payudara, menggunakan dataset dari Bioimaging 2015 breast histology classification challenge. Dataset tersebut berisi citra resolusi tinggi (2040 x 1536 pixel), dengan pembesaran 200X, setiap citra dilabeli dengan salah satu dari empat label yaitu, normal tissue, benign lesion, in situ carcinoma, dan invasive carcinoma yang dilakukan oleh dua ahli patologi. Dataset terdiri dari set data latih 249 citra, dan 20 set data tes terpisah, serta data test tambahan sejumlah 16 citra. Metode CNN digunakan untuk melakukan klasifikasi dengan arsitektur, lima convolutional layer yang diikuti max polling dan tiga fully connected layer. Dari hasil penelitian tersebut akurasi yang didapatkan adalah 77.8% untuk empat kelas, dan 83.3% untuk carcinoma/non-carcinoma (8).

Liu, et. al., menggunakan fully-connected layer first convolutional neural network (FCLF-CNN) untuk melakukan klasifikasi kanker payudara. Dataset WBCD dan WDBC yang digunakan merupakan sample citra aspirasi jarum halus jaringan payudara manusia, dimana WBCD memiliki 596 sample dan WDBC memiliki 699 sample. FCLF-CNN yang digunakan pada penelitian ini memberikan hasil yang lebih baik dari pada metode CNN biasa, dan ketika menggunakan ensemble FCLF-CNN memberikan hasil yang lebih baik dari pada single FCLF-CNN dengan akurasi 98.71% pada dataset WBCD dan akurasi 99.28% pada dataset WDBC (9).

Angel Cruz-Roa et al. dan Andrew Janowczyk et al. melakukan penelitian klasifikasi kanker payudara dengan dataset yang digunakan berasal dari 162 pasien berisi citra whole slide image (WSI). Angel Cruz-Roa et al. menggunakan CNN untuk melakukan klasifikasi dengan ukuran citra 100x100 pixel, memberikan hasil f1-score 0.7180 dan balanced accuracy 0.8423. penelitian yang dilakukan oleh Andrew Janowczyk et al. menggunakan AlexNet dengan segmentasi pada dataset, ukuran citra di-resize menjadi 32x32 memberikan hasil yang lebih baik dengan f1-score 0.7648 dan balanced accuracy 0.8468 (10,11).

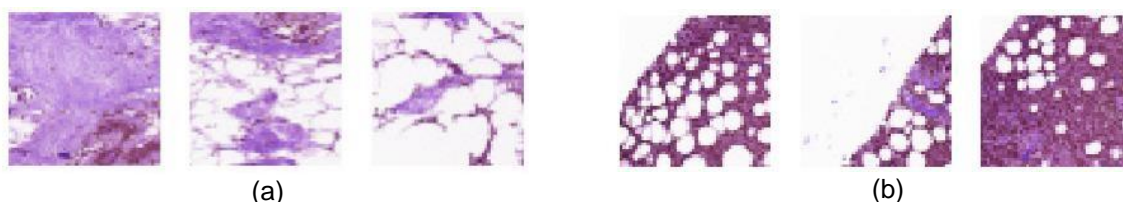
Aiza M. Romano et. al. pada tahun 2019 melakukan penelitian klasifikasi kanker payudara dengan metode deep learning menggunakan CNN yang telah ditingkatkan performanya. Menggunakan dataset yang sama dengan penelitian (10,11). Dataset yang digunakan berisi 277,524 citra dari 162 pasien, yang terdiri dari 198,738 Invasive Ductal Carcinoma (IDC) negatif dan 78,786 citra IDC positif. Dari keseluruhan dataset dilakukan under sampling terhadap kelas mayoritas untuk mengatasi kelas yang tidak seimbang, yang kemudian total kedua kelas dipisah dengan rasio 80:20 menjadi 40672 data train dan 10186 data test, yang masing masing terdiri dari dua kategori, yaitu IDC positif dan IDC negatif. Dengan model CNN yang digunakan, mendapatkan performa F1-score 0.8528 dan akurasi seimbang 0.8541 (12).

Pada penelitian yang dilakukan oleh Aiza M. Romano et. al. terlihat bahwa akurasi seimbang dan f1-score masih terdapat ruang untuk peningkatan. Melihat peningkatan yang bisa didapatkan pada penelitian Liu et. al. dengan mengaplikasikan ensemble pada model CNN. Pada penelitian ini akan mengusulkan metode ensemble dengan model CNN untuk mendapatkan peningkatan performa dibandingkan metode pada penelitian Aiza M. Romano et. al.

2. Metode Penelitian

2.1 Pengambilan Dataset

Dataset yang digunakan pada penelitian ini diperoleh dari Kaggle dengan judul Breast Histopathology Images, berisi total 277,524 citra dengan ukuran 50x50 dengan format png. 198,738 citra untuk Invasive Ductal Carcinoma (IDC) negatif dan 78,786 untuk IDC positif. Sumber dataset dapat diakses melalui website <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>. Contoh sampel dapat dilihat pada Gambar 1.



Gambar 1. Contoh sample IDC (a) IDC negatif, (b) IDC positif

2.2 Data preprocessing dan Data Splitting

Dataset yang didapatkan berisi citra-citra histologi dari banyak id pasien, citra-citra tersebut berada dalam masing-masing folder sebanyak 279 folder dengan nama sesuai id pasien dan di dalamnya terbagi menjadi dua kelas, yaitu kelas 0 dan 1.

Pada dataset tersebut terdapat permasalahan jumlah dataset yang tidak seimbang antara kedua kelas, tantangan dalam mempelajari data tidak seimbang adalah kecenderungan terjadinya bias, dikarenakan pengklasifikasi lebih banyak melihat data mayoritas sehingga ada kemungkinan data minoritas dianggap sebagai noise (13). Untuk mengatasi itu dilakukan pemilihan data secara acak hingga masing masing kelas memiliki jumlah 25420 data, sehingga total dataset yang digunakan sebanyak 50840.

Dataset yang telah diproses tersebut kemudian dilakukan splitting dataset, dengan rasio data train 80% dan data test 20%. Hasil splitting dataset dapat dilihat pada Tabel 1 berikut.

Tabel 1. Hasil splitting dataset

Label	Kelas	Jumlah sample data train	Jumlah sample data test	Total
1	IDC positif	20.336	5.084	25420
0	IDC negatif	20.336	5.084	25420
				50840

2.3 Data Augmentasi

Data augmentasi umumnya digunakan sebagai alternatif untuk memperbanyak data secara artifisial berdasarkan sample yang ada, dapat juga digunakan untuk mengatasi invarian pada data dengan beberapa teknik transformasi seperti, rotasi, flip, cropping, dll (14). Proses augmentasi dilakukan dengan bantuan library Keras ImageDataGenerator. Citra-citra tersebut akan melalui serangkaian transformasi acak, parameter yang digunakan pada proses augmentasi dapat dilihat pada Tabel 2.

Tabel 2. Parameter ImageDataGenerator

Teknik Augmentasi	Parameter
rescale	1.0/255
shear_range	0.2
zoom_range	0.2
horizontal_flip	True
vertical_flip	True
rotation_range	20
width_shift_range	0.2
height_shift_range	0.2

2.4 Model

Convolutional neural network (CNN), sering juga disebut ConvNet digambarkan sebagai konsep detector fitur dengan cara kerja yang terinspirasi dari sistem biologi. Mampu mempelajari fitur yang sangat abstrak dan dapat mengidentifikasi objek secara efisien (15). CNN adalah kombinasi dari algoritma deep learning dan artificial neural network. Umumnya terdiri dari tiga bagian, pertama input layer yang berisi citra asli yang belum diproses, kedua ada output layer yang merupakan hasil dari mengklasifikasikan fitur, dan yang ke tiga adalah hidden layer yang merupakan layer neuron dengan struktur nonlinear multi-layer yang kompleks (16).

Penggunaan deep learning memiliki kelebihan dalam menemukan struktur yang rumit pada data yang memiliki dimensional tinggi. Memiliki arsitektur yang dibangun dari banyak lapisan modul sederhana yang disusun secara hirarkis. Modul-modul tersebut bertugas untuk melakukan perhitungan representasi dari lapisan lapisan sebelumnya, serta bertanggung jawab meningkatkan invarian dan kepekaan representasi dengan mengubah input dari lapisan sebelumnya (17). Deep learning secara umum terbagi menjadi tiga kategori, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Pada penelitian ini yang digunakan adalah supervised learning dan convolutional neural network adalah salah satu contohnya (18).

Pada penelitian ini model CNN yang digunakan mengimplementasikan transfer learning dan juga fine tuning. Transfer learning pada dasarnya merupakan proses transfer dan pemanfaatan kembali pengetahuan dari suatu bidang ke bidang lain yang berhubungan. Karena pada dasarnya informasi dasar yang diekstrak dari suatu data memiliki kemiripan sekalipun dibidang yang berbeda, misalnya, pada citra motor dan tumbuhan yang sama sama memiliki informasi warna rgb. Transfer learning umumnya menggunakan model deep learning yang telah dilatih dengan dataset besar, untuk kemudian struktur dan parameter model tersebut ditransfer ke tugas baru sebagai feature extraction dan dilatih dengan dataset baru, dengan tetap menggunakan bobot aslinya sebagai bobot awal yang disebut dengan fine tuning (19–21).

Model MobileNet, MobileNetV2 dan VGG16 digunakan pada penelitian ini, dengan mengimplementasikan transfer learning. Pada kedua model tersebut, classifier asli tidak digunakan dan ditambahkan classifier baru yang akan dilatih untuk mengklasifikasikan citra kanker payudara pada penelitian ini, sehingga hanya menyisakan feature extraction saja dari arsitektur asli model. Penambahan fully connected layer atau classifier dapat dilihat pada table 3.

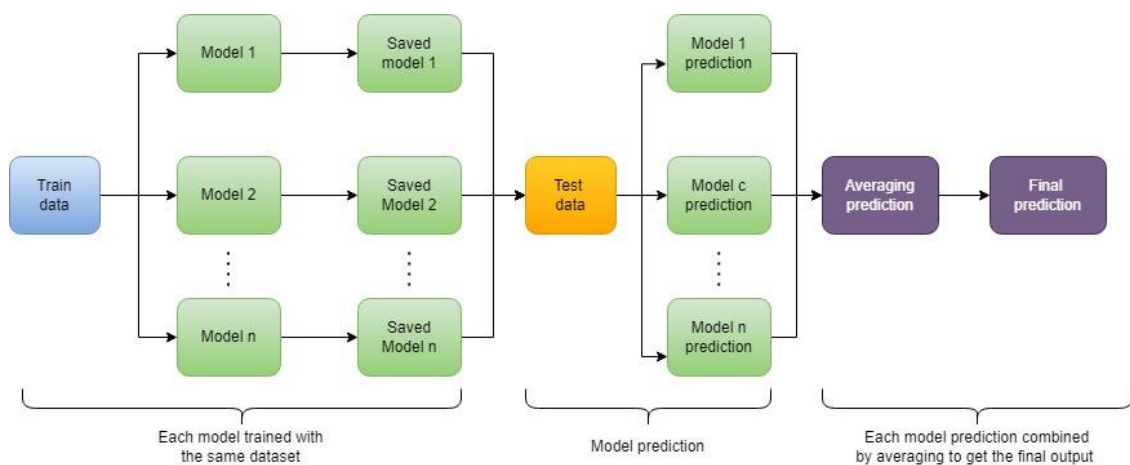
Tabel 3. Fully connected layer

Layer	Parameter
Flaten	-
Dense	128, relu
Dropout	0.5
Dense	2, softmax

2.5 Ensemble

Ensemble menggabungkan beberapa model, sehingga ensemble juga bisa disebut system multiclassifier. Dengan menggabungkan beberapa pembelajaran atau model, ensemble bisa memberikan performa yang lebih baik dibanding pembelajaran single model dan juga meminimalisir kemungkinan pemilihan dari hasil yang buruk (22,23). Akurasi ensemble sendiri dipengaruhi oleh jumlah model, diversity, dan akurasi dari tiap model yang digunakan(24).

Metode ensemble yang digunakan pada penelitian ini adalah averaging. Model yang sudah disimpan sebelumnya akan digabung menjadi satu pengklasifikasi untuk melakukan prediksi, yang kemudian hasil prediksi tiap model terhadap dua kelas label akan di rata rata untuk menjadi prediksi akhir.



Gambar 2. Contoh skema/diagram ensemble averaging

2.6 Skenario Pengujian

Pada tahap ini dilakukan pengujian model usulan MobileNet, MobileNetV2 dan VGG16 terhadap dataset yang telah melalui proses augmentasi. Ketiga model tersebut menggunakan parameter yang sama, pengujian dilakukan dengan epoch 64, steps_per_epoch 128, batch size 64, learning rate 0.001 dan 0.0001 (fine tuning), optimizer Adam, dan menggunakan callbacks ModelCheckpoint dengan memonitor val_accuracy.

Ketiga selain dilakukan transfer learning juga dilakukan fine tuning, yaitu dengan melatih ketiga model pada kondisi layer trainable dari base model dibekukan, menyisakan classifier yang kita tambahkan untuk dilatih dengan dataset baru. Proses training pertama menggunakan parameter yang telah disebutkan namun untuk learning rate menggunakan 0.001. setelahnya, proses fine tuning, yaitu training model kedua kalinya namun dengan kondisi layer trainable tidak dibekukan dan menggunakan learning rate 0.0001.

Metode averaging digunakan pada proses ensemble untuk menentukan hasil prediksi. Ketiga model sebelumnya melakukan prediksi terhadap dataset, kemudian hasil output ketiga model akan di rata rata dan menjadi prediksi akhir.

3. Hasil Penelitian dan Pembahasan

3.1 Pengujian Model

Pada proses ini dilakukan pengujian terhadap Ketiga model dengan data set yang telah diaugmentasi. Pengujian dilakukan dengan parameter yang sama pada ketiga model. Training pertama dilakukan dengan learning rate 0.001 terlebih dahulu dengan kondisi trainable layer "False". Setelah proses training pertama selesai, training kedua dilakukan dengan learning rate 0.0001 untuk proses fine tuning dengan kondisi trainable layer "True".

3.2 Ensemble

Dari ketiga model yang telah melalui proses training tersebut, dilakukan proses ensemble, dengan menghitung rata rata dari hasil klasifikasi tiap model pada tiap citra yang diprediksi.

3.3 Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui performa yang didapatkan dari ketiga model, evaluasi yang dilakukan meliputi roc-auc score, classification report, grafik akurasi dan loss, confusion matrix serta balanced accuracy dan f1-score. Evaluasi pertama dilakukan terhadap roc-auc score dari ketiga model, hal ini untuk melihat seberapa baik model dalam melakukan klasifikasi binary, yaitu memisahkan class 0 sebagai 0 dengan class 1 sebagai 1. Hasil roc-auc score dapat dilihat pada Tabel 4.

Tabel 4. Hasil roc-auc score

Model	Roc-auc score
MobileNet	0.9321
MobileNetV2	0.9303
VGG16	0.9353

Selanjutnya, membuat classification report untuk melihat seberapa baik kualitas prediksi yang dilakukan oleh model, dan dari laporan ini kita dapat memahami performa dari tiap model. Hasil classification report dapat dilihat pada Tabel 5.

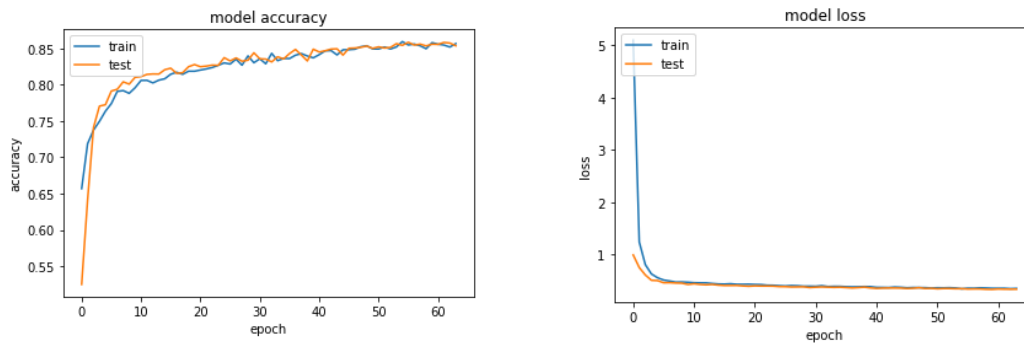
Tabel 5. Hasil classification report

Model	Class	Precision	Recall	F1-score	Support
MobileNet	No_idc	0.84	0.88	0.86	5084
	Has_idc	0.87	0.83	0.85	5084
	Accuracy			0.85	10168
	Macro avg	0.86	0.86	0.86	10168
	Weighted avg	0.86	0.86	0.86	10168
MobileNetV2	No_idc	0.81	0.91	0.85	5084
	Has_idc	0.89	0.79	0.84	5084
	Accuracy			0.85	10168

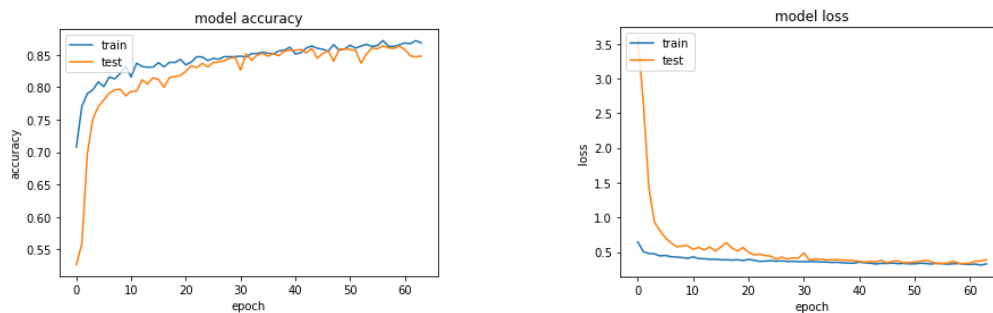
	Macro avg	0.85	0.85	0.85	10168
	Weighted avg	0.85	0.85	0.85	10168
VGG16	No_idc	0.87	0.84	0.86	5084
	Has_idc	0.85	0.88	0.86	5084
	Accuracy			0.86	10168
	Macro avg	0.86	0.86	0.86	10168
	Weighted avg	0.86	0.86	0.86	10168

Dari Tabel 5 kita dapat memahami beberapa hal. Precision menunjukkan kemampuan classifier untuk memberikan prediksi secara tepat, MobileNetV2 dapat memprediksi label “Has_idc” lebih baik dari dua model lainnya dan VGG16 memprediksi label “No_idc” lebih baik dari dua model lainnya. Recall menunjukkan berapa persen prediksi positif yang didapatkan, VGG16 mendapat skor tertinggi pada label “Has_idc”, dan MobileNetV2 pada label “No-idc”. F1-score menunjukkan berapa persen prediksi positif yang ditebak dengan benar, model VGG16 mendapat skor paling baik dibandingkan kedua model lainnya. dari ketiga hal tersebut dapat dikatakan bahwa model VGG16 memiliki kualitas hasil yang lebih baik.

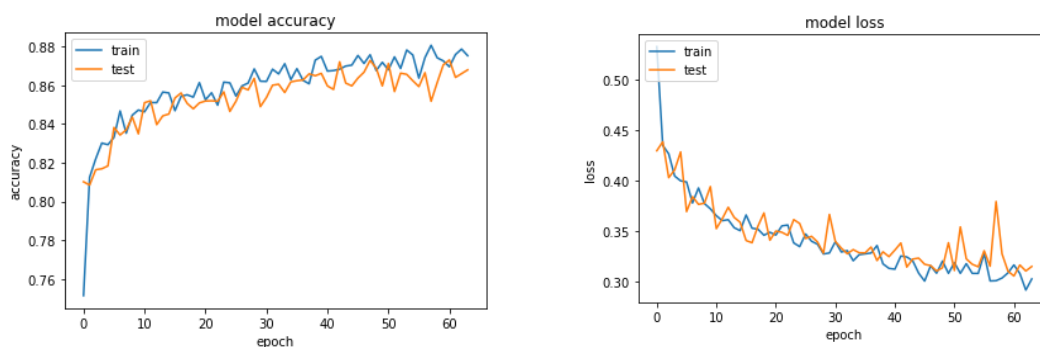
Dilakukan plot grafik akurasi dan loss dari setiap epoch, untuk melihat keadaan model pada saat training. hasil grafik dapat dilihat pada Gambar 3, Gambar 4, dan Gambar 5.



Gambar 3. Grafik MobileNet

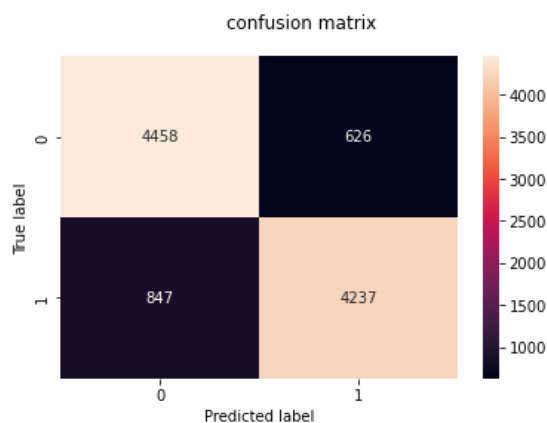


Gambar 4. Grafik MobileNetV2

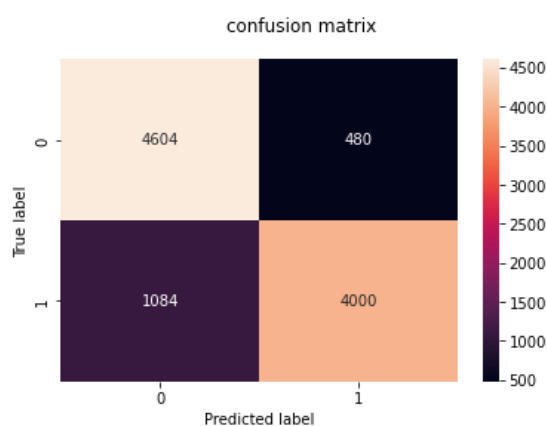


Gambar 5. Grafik VGG16

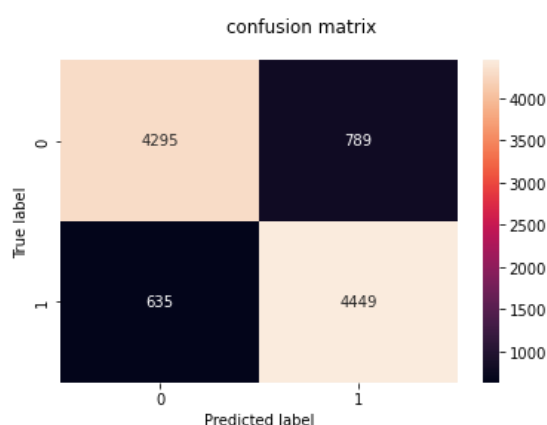
Dari grafik plot ketiga model terlihat bahwa model tidak mengalami overfitting dan underfitting, terutama pada model MobileNet. Setelahnya, dilakukan plot untuk confusion matrix, dari confusion matrix kita dapat memahami hasil prediksi dari tiap model secara detail, berapa yang positif dan negative dari kedua kelas atau label. plot confusion matrix dapat dilihat pada Gambar 6, Gambar 7, dan Gambar 8.



Gambar 6. MobileNet



Gambar 7. MobileNetV2



Gambar 8. VGG16

Dari confusion matrix tersebut, tersebut dilakukan perhitungan f1-score dan balanced accuracy dari kedua model. Hasil f1-score dan balanced accuracy dapat dilihat pada Tabel 6.

Tabel 6. Hasil roc-auc score

Hasil	Balanced accuracy	F1-score
MobileNet	0.8551	0.8519
MobileNetV2	0.8461	0.8364
VGG16	0.8599	0.8620

Berdasarkan hasil evaluasi pada Tabel 6, terlihat bahwa model VGG16 memberikan hasil yang baik dibandingkan model lain dari evaluasi roc-auc score, balanced accuracy, dan f1-score paling tinggi.

3.4 Evaluasi Ensemble

Pada tahap evaluasi hasil ensemble, dilakukan hal yang sama dengan evaluasi model, hanya saja tanpa grafik loss dan akurasi. Proses ensemble dilakukan dengan semua kombinasi dari ketiga model untuk mendapatkan kombinasi terbaik. Pertama dengan melakukan evaluasi terhadap roc-auc score yang dapat dilihat pada Tabel 7.

Tabel 7. Hasil roc-auc score

Kombinasi Model	ROC-AUC score
VGG16 + MobileNet	0.9409
VGG16 + MobileNetV2	0.9404
MobileNet + MobileNetV2	0.9352
VGG16 + MobileNet + MobileNetV2	0.9413

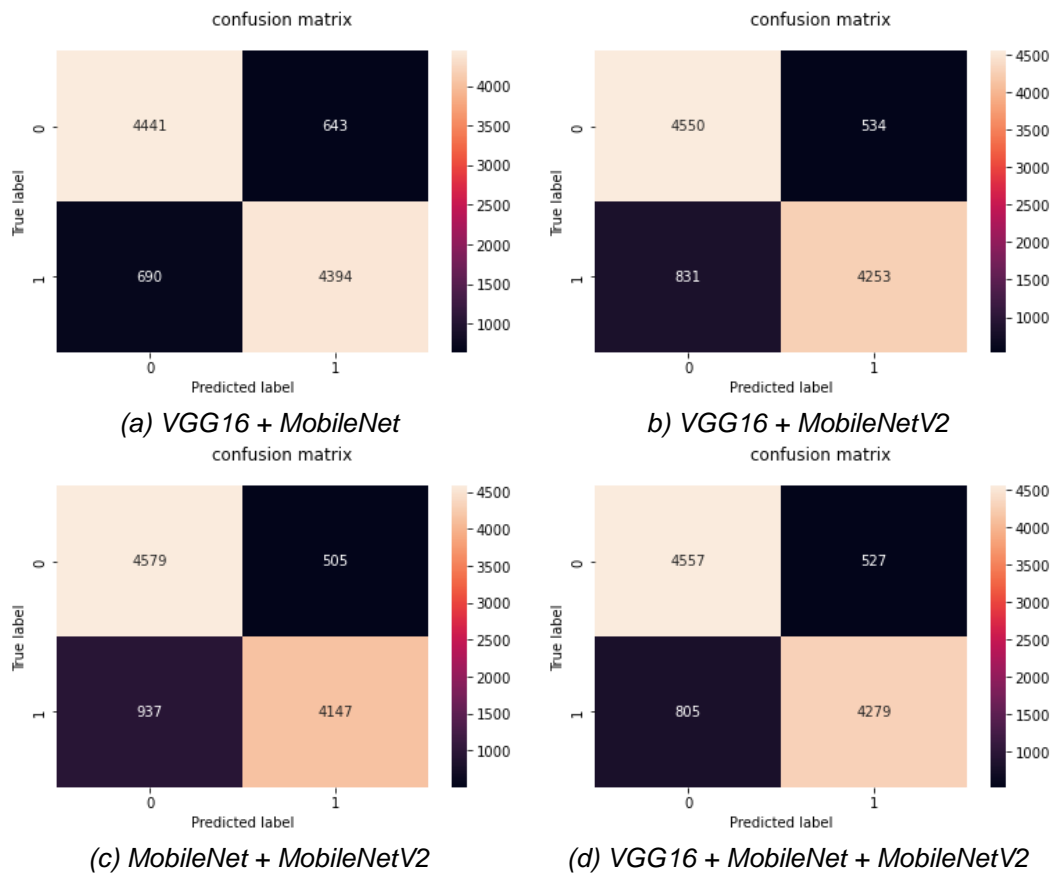
Dari hasil roc-auc score dapat dilihat bahwa kombinasi dengan tiga model memberikan roc-auc paling baik, diikuti oleh kombinasi VGG16 + MobileNet. Kemudian evaluasi classification report dapat dilihat pada Tabel 8.

Tabel 8. Hasil roc-auc score

Model	Class	Precision	Recall	F1-score	Support
VGG16 + MobileNet	No_idc	0.87	0.87	0.87	5084
	Has_idc	0.87	0.86	0.87	5084
	Accuracy			0.87	10168
	Macro avg	0.87	0.87	0.87	10168
	Weighted avg	0.87	0.87	0.87	10168
VGG16 + MobileNetV2	No_idc	0.85	0.89	0.87	5084
	Has_idc	0.89	0.84	0.86	5084
	Accuracy			0.87	10168
	Macro avg	0.87	0.87	0.87	10168
	Weighted avg	0.87	0.87	0.87	10168
MobileNet + MobileNetV2	No_idc	0.83	0.90	0.86	5084
	Has_idc	0.89	0.82	0.85	5084
	Accuracy			0.86	10168
	Macro avg	0.86	0.86	0.86	10168
	Weighted avg	0.86	0.86	0.86	10168
VGG16 + MobileNet + MobileNetV2	No_idc	0.85	0.90	0.87	5084
	Has_idc	0.89	0.84	0.87	5084
	Accuracy			0.87	10168
	Macro avg	0.87	0.87	0.87	10168
	Weighted avg	0.87	0.87	0.87	10168

Selanjutnya membuat Grafik confusion matrix ensemble, dan dilakukan perhitungan balanced accuracy dan f1-score dari masing-masing kombinasi ensemble yang didapatkan dari

confusion matrix. Hasil confusion matrix dapat dilihat pada gambar 9 dan hasil perhitungan dapat dilihat pada Tabel 9.



Gambar 9. VGG16

Tabel 9. Hasil perhitungan balanced accuracy dan f1-score

Hasil	Balanced accuracy	F1-score
VGG16 + MobileNet	0.8689	0.8682
VGG16 + MobileNetV2	0.8657	0.8617
MobileNet + MobileNetV2	0.8581	0.8518
VGG16 + MobileNet + MobileNetV2	0.8690	0.8653

Dari hasil evaluasi ensemble didapatkan kombinasi terbaik adalah VGG16 + MobileNet dan VGG16 + MobileNet + MobileNetV2. Kedua kombinasi tersebut lebih baik dari dua kombinasi lainnya, namun pada kedua kombinasi terbaik, VGG16 + MobileNet lebih unggul pada f1-score dan VGG16 + MobileNet + MobileNetV2 lebih unggul sedikit pada balanced accuracy.

3.5 Perbandingan Hasil

Pada Tahap ini akan dibahas mengenai perbandingan hasil yang didapatkan pada penelitian ini baik model dan ensemble dengan penelitian sebelumnya. Detail mengenai tabel perbandingan dapat dilihat pada Tabel 10.

Tabel 2. Perbandingan hasil seluruh model

Metode	Balanced accuracy	F1-score
CNN (10)	0.8423	0.7180
Alexnet with Resize (11)	0.8468	0.7648
Penelitian rujukan (CNN) (12)	0.8528	0.8541
MobileNet	0.8551	0.8519

MobileNetV2	0.8461	0.8364
VGG16	0.8599	0.8620
VGG16 + MobileNet	0.8689	0.8682
VGG16 + MobileNetV2	0.8657	0.8617
MobileNet + MobileNetV2	0.8581	0.8518
VGG16 + MobileNet + MobileNetV2	0.8690	0.8653

Penelitian sebelumnya yang dilakukan oleh (10–12) menggunakan dataset sama dengan penelitian ini. Hasil performa model VGG16 yang didapatkan pada penelitian ini dapat dikatakan lebih baik dari penelitian sebelumnya yang menggunakan kustom CNN dan AlexNet.

Dengan penggunaan transfer learning, penelitian ini diuntungkan pada pembuatan model yang hanya perlu menggunakan arsitektur pre-trained model dengan penambahan classifier dibandingkan penelitian (10,12) yang menggunakan custom CNN. Dibandingkan penelitian (11) yang menggunakan tambahan proses segmentation secara manual, yang tentunya akan memakan waktu dan juga kemungkinan terjadinya perbedaan keakuratan pada pengamatan (25), pendekatan learn-from-data pada penelitian ini dan juga pada penelitian (12) tentunya membuat penelitian menjadi lebih sederhana dan mudah. Selain itu juga fine tuning yang dilakukan membuat pengetahuan atau bobot dari pre-trained model tetap dapat dimanfaatkan bersamaan dengan pengetahuan atau bobot dari dataset baru. Metode ensemble dengan kombinasi VGG16 yang dilakukan juga memberikan peningkatan performa dibandingkan penggunaan single model VGG16 terutama pada skor balanced accuracy.

Referensi

- [1] Man R, Yang P, Xu B. Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks. *IEEE Access*. 2020;8:155362–77.
- [2] Press Release N° 292 [Internet]. 2020 [cited 2021 Jan 31]. Available from: <https://gco.iarc.fr/>,
- [3] Bardou D, Zhang K, Ahmad SM. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. *IEEE Access*. 2018 May 1;6:24680–93.
- [4] Li Y, Wu J, Wu Q. Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning. *IEEE Access*. 2019;7:21400–8.
- [5] Wilson ML, Fleming KA, Kuti MA, Looi LM, Lago N, Ru K. Access to pathology and laboratory medicine services: a crucial gap [Internet]. Vol. 391, *The Lancet*. Lancet Publishing Group; 2018 [cited 2021 Feb 1]. p. 1927–38. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0140673618304586>
- [6] Abu MA, Indra NH, Rahman AHA, Sapiee NA, Ahmad I. A study on image classification based on deep learning and tensorflow. *Int J Eng Res Technol*. 2019;12(4):563–9.
- [7] Lei X, Pan H, Huang X. A dilated cnn model for image classification. *IEEE Access*. 2019;7:124087–95.
- [8] Araujo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of breast cancer histology images using convolutional neural networks. Sapino A, editor. *PLoS One* [Internet]. 2017 Jun 1 [cited 2021 Feb 1];12(6):e0177544. Available from: <https://dx.plos.org/10.1371/journal.pone.0177544>
- [9] Liu K, Kang G, Zhang N, Hou B. Breast Cancer Classification Based on Fully-Connected Layer First Convolutional Neural Networks. *IEEE Access*. 2018 Mar 19;6:23722–32.
- [10] Cruz-Roa A, Basavanhally A, González F, Gilmore H, Feldman M, Ganesan S, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. *Med Imaging 2014 Digit Pathol*. 2014;9041(216):904103.
- [11] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016;7(1):904103.
- [12] Romano AM, Hernandez AA. Enhanced Deep Learning Approach for Predicting Invasive Ductal Carcinoma from Histopathology Images. 2019 2nd Int Conf Artif Intell Big Data, ICAIBD 2019. 2019;142–8.
- [13] Khushi M, Shaukat K, Alam TM, Hameed IA, Uddin S, Luo S, et al. A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access*. 2021;9:109960–75.
- [14] Zhang X, Wang Z, Liu D, Lin Q, Ling Q. Deep Adversarial Data Augmentation for Extremely Low Data Regimes. *IEEE Trans Circuits Syst Video Technol*. 2021;31(1):15–28.

- [15] Indolia S, Goswami AK, Mishra SP, Asopa P. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. In: *Procedia Computer Science* [Internet]. Elsevier B.V.; 2018 [cited 2021 Feb 1]. p. 679–88. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1877050918308019>
- [16] Tian Y. Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access*. 2020;8:125731–44.
- [17] Zhang S, Bamakan SMH, Qu Q, Li S. Learning for Personalized Medicine: A Comprehensive Review from a Deep Learning Perspective. *IEEE Rev Biomed Eng*. 2018;12(XX):194–208.
- [18] Zhu T, Li K, Herrero P, Georgiou P. Deep Learning for Diabetes: A Systematic Review. *IEEE J Biomed Heal Informatics*. 2021;25(7):2744–57.
- [19] Wang W, Wang C, Wang Z, Yuan M, Luo X, Kurths J, et al. Abnormal detection technology of industrial control system based on transfer learning. *Appl Math Comput*. 2022;412(6):821–32.
- [20] Ahmad M, Abdullah M, Moon H, Han D. Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks with Stepwise Transfer Learning. *IEEE Access*. 2021;9:140565–80.
- [21] Do S, Song KD, Chung JW. Basics of deep learning: A radiologist's guide to understanding published radiology articles on deep learning. *Korean J Radiol*. 2020;21(1):33–41.
- [22] Wang Z, Chu R, Zhang M, Wang X, Luan S. An Improved Selective Ensemble Learning Method for Highway Traffic Flow State Identification. *IEEE Access*. 2020;8:212623–34.
- [23] Xue D, Zhou X, Li C, Yao Y, Rahaman MM, Zhang J, et al. An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification. *IEEE Access*. 2020;8:104603–18.
- [24] Yang J, Wang F. Auto-Ensemble: An Adaptive Learning Rate Scheduling Based Deep Learning Model Ensembling. *IEEE Access*. 2020;8:217499–509.
- [25] Van Opbroek A, Achterberg HC, Vernooij MW, De Bruijne M. Transfer learning for image segmentation by combining image weighting and kernel learning. *IEEE Trans Med Imaging*. 2019;38(1):213–24.

