

# Klasifikasi Stadium Fibrosis Hati pada Citra USG Menggunakan EfficientNet-B0

Ananda Putra Wahyu Riyanto<sup>1</sup>, Faisal Muttaqin<sup>2</sup>, Eva Yulia Puspaningrum<sup>3</sup>

<sup>1,2,3</sup> Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur

<sup>1</sup>[22081010038@student.upnjatim.ac.id](mailto:22081010038@student.upnjatim.ac.id)

<sup>2</sup>[faisalmuttaqin.if@upnjatim.ac.id](mailto:faisalmuttaqin.if@upnjatim.ac.id)

<sup>3</sup>[evapuspaningrum.if@upnjatim.ac.id](mailto:evapuspaningrum.if@upnjatim.ac.id)

\*Corresponding author email: [22081010038@student.upnjatim.ac.id](mailto:22081010038@student.upnjatim.ac.id)

**Abstrak**— Penilaian stadium fibrosis hati penting dalam praktik klinis karena berhubungan dengan risiko komplikasi dan penentuan terapi. Namun, pembacaan citra ultrasonografi (USG) masih sangat dipengaruhi pengalaman operator. Penelitian ini menilai kemampuan EfficientNet-B0 dalam mengklasifikasikan lima stadium fibrosis hati (F0–F4) menggunakan 6.323 citra USG. Data dibagi secara *stratified* menjadi data latih, validasi, dan uji dengan rasio 70:15:15. Sebelum pelatihan, seluruh citra diubah ke format RGB berukuran 224×224 piksel dan pada data latih diterapkan augmentasi berupa pembalikan horizontal serta rotasi kecil. Model dilatih dengan pendekatan *transfer learning*, fungsi *loss* berbobot kelas, *optimizer* AdamW, dan *scheduler* ReduceLROnPlateau. Penilaian kinerja dilakukan menggunakan *accuracy*, *macro-F1*, *macro-AUROC*, *confusion matrix*, serta evaluasi kalibrasi melalui *reliability diagram* dan Expected Calibration Error (ECE). Hasil pengujian menunjukkan bahwa model mencapai *accuracy* 0,9779, *macro-F1* 0,9671, dan *macro-AUROC* 0,9983, dengan 21 kesalahan klasifikasi dari 949 data uji. Model sangat baik dalam mengenali kelas ekstrem, tetapi masih mengalami kesulitan pada kelas menengah yang memiliki tampilan visual lebih mirip. Dari sisi probabilitas prediksi, nilai ECE 0,0161 menunjukkan bahwa tingkat keyakinan model cukup sesuai dengan akurasi aktual. Temuan ini menunjukkan bahwa EfficientNet-B0 memiliki potensi untuk digunakan sebagai dasar sistem bantu baca USG pada *staging* fibrosis hati, meskipun uji pada data eksternal tetap diperlukan sebelum penerapan klinis.

**Kata Kunci**— fibrosis hati, ultrasonografi, *deep learning*, EfficientNet-B0, kalibrasi probabilitas

## I. PENDAHULUAN

Fibrosis hati adalah proses yang berlangsung bertahap dan dapat berkembang menjadi sirosis, hipertensi portal, hingga karsinoma hepatoselular. Karena itu, penentuan stadium fibrosis menjadi bagian penting dalam menilai risiko pasien, memilih terapi, dan memantau perkembangan penyakit [1]. Penilaian yang akurat juga dibutuhkan karena berhubungan langsung dengan prognosis dan tindak lanjut klinis [2]. Dalam praktik sehari-hari, ultrasonografi (USG) masih menjadi salah satu pemeriksaan awal yang paling sering digunakan karena mudah diperoleh, relatif murah, dan aman [3]. Walaupun elastografi memberikan pendekatan yang lebih objektif untuk menilai kekakuan jaringan, citra B-mode tetap banyak dipakai karena alatnya lebih tersedia dan proses pemeriksaannya lebih sederhana [4]. Kendalanya, hasil interpretasi USG masih dipengaruhi operator, kualitas

pengambilan citra, kondisi pasien, dan perbedaan perangkat, sehingga penilaian antar pemeriksa belum selalu konsisten [3]. Pemanfaatan *deep learning* dalam citra medis terus berkembang dan telah menunjukkan hasil yang baik pada berbagai tugas klasifikasi maupun segmentasi [5]. Pada citra USG, pendekatan ini menarik karena mampu membantu pembacaan citra yang umumnya memiliki *speckle noise*, kontras rendah, dan variasi akuisisi [6]. Sejumlah penelitian juga melaporkan bahwa model berbasis *convolutional neural network* (CNN) dapat digunakan untuk mengenali stadium fibrosis hati dari citra USG dengan hasil yang menjanjikan [7], [8].

Salah satu arsitektur CNN yang menarik adalah EfficientNet-B0. Model ini dirancang untuk memperoleh keseimbangan antara ketelitian dan efisiensi komputasi melalui strategi *compound scaling* [9]. Karakter tersebut penting untuk pengembangan sistem bantu baca klinis, karena model yang tidak terlalu berat lebih realistis diterapkan. Selain itu, penggunaan *transfer learning* dari bobot pralatih juga umum dilakukan pada data medis untuk membantu proses pelatihan menjadi lebih stabil [10].

Meski demikian, evaluasi model medis tidak cukup dilihat dari akurasi saja. Distribusi kelas yang tidak seimbang membuat metrik seperti *macro-F1* lebih relevan untuk melihat performa antar kelas secara lebih adil [11]. Selain itu, nilai probabilitas keluaran model juga perlu diperiksa, karena *confidence* yang tinggi tidak selalu berarti prediksi benar. Oleh sebab itu, analisis kalibrasi menggunakan *reliability diagram* dan ECE menjadi penting [12], [13]. Dalam kasus fibrosis, hal ini makin relevan karena label F0 hingga F4 memiliki urutan klinis, sehingga kesalahan pada kelas yang berdekatan dan berjauhan tidak memiliki dampak yang sama [14].

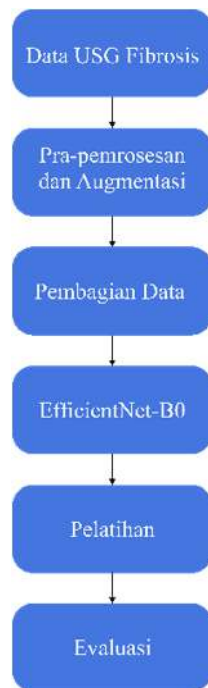
Berdasarkan hal tersebut, penelitian ini mengevaluasi EfficientNet-B0 untuk klasifikasi lima stadium fibrosis hati pada citra USG. Model dilatih dengan *transfer learning* dan dinilai menggunakan *accuracy*, *macro-F1*, *macro-AUROC*, *confusion matrix*, *reliability diagram*, dan ECE untuk melihat ketepatan klasifikasi sekaligus keandalan probabilitas prediksi.

## II. METODE PENELITIAN

### A. Desain Penelitian

Penelitian ini merupakan eksperimen klasifikasi multikelas untuk mengenali lima stadium fibrosis hati, yaitu F0 - F4, dari

citra USG. Model yang digunakan adalah EfficientNet-B0, yaitu CNN yang dipilih karena relatif ringan tetapi tetap mampu mengekstraksi fitur citra dengan baik. Model diawali dengan bobot pralatih melalui *transfer learning* agar proses belajar lebih stabil pada data medis. Tahapan penelitian mencakup pengumpulan data, pra-pemrosesan, augmentasi, pembagian data, pelatihan model, dan evaluasi hasil, sebagaimana ditunjukkan pada Gbr. 1.



Gbr 1. Diagram Alur Penelitian

### B. Dataset dan Pembagian Data

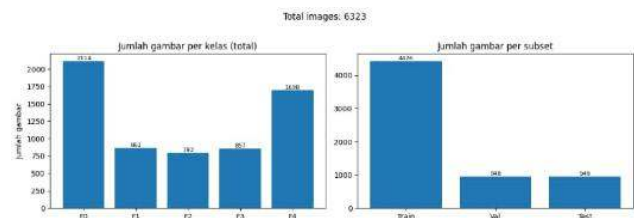
Dataset yang digunakan pada penelitian ini diperoleh dari platform Kaggle dengan nama “Liver Histopathology (Fibrosis) Ultrasound Images”. Dataset tersebut berupa citra ultrasonografi (USG) hati yang dikelompokkan ke dalam lima kelas stadium fibrosis, yaitu F0, F1, F2, F3, dan F4. Pelabelan dilakukan mengikuti struktur folder pada dataset sehingga setiap citra memperoleh label sesuai kelasnya. Secara keseluruhan terdapat 6.323 citra, dengan distribusi F0 = 2.114, F1 = 861, F2 = 793, F3 = 857, dan F4 = 1.698. Rincian distribusi jumlah citra pada setiap kelas dan pembagian subset disajikan pada Tabel I.

TABEL I  
DISTRIBUSI DATASET DAN PEMBAGIAN SUBSET

| Kelas | Jumlah Total | Train | Validation | Test |
|-------|--------------|-------|------------|------|
| F0    | 2114         | ...   | ...        | ...  |

|       |      |      |     |     |
|-------|------|------|-----|-----|
| F1    | 861  | ...  | ... | ... |
| F2    | 793  | ...  | ... | ... |
| F3    | 857  | ...  | ... | ... |
| F4    | 1698 | ...  | ... | ... |
| Total | 6323 | 4426 | 948 | 949 |

Setelah proses pengelompokan, dataset dibagi menjadi tiga bagian subset, yaitu data latih, data validasi, dan data uji dengan rasio 70:15:15 menggunakan *stratified split* agar proporsi masing-masing kelas tetap terjaga pada setiap bagian. Hasil split menghasilkan 4.426 citra pada data latih, 948 citra pada data validasi, dan 949 citra pada data uji. Visualisasi distribusi jumlah citra per kelas dan pembagian data ditunjukkan pada Gbr. 2.



Gbr. 2 Distribusi jumlah citra per kelas dan pembagian data

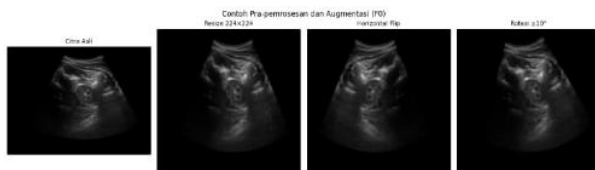
Komposisi tersebut menunjukkan bahwa dataset bersifat tidak seimbang, karena kelas F0 dan F4 memiliki jumlah citra yang lebih besar dibandingkan kelas menengah F1-F3. Kondisi ini perlu diperhatikan karena berpotensi menyebabkan model lebih bias terhadap kelas mayoritas. Oleh karena itu, ketidakseimbangan distribusi kelas ini menjadi salah satu dasar penggunaan strategi pembobotan kelas pada tahap pelatihan. Data latih digunakan untuk memperbarui parameter model, data validasi digunakan untuk memantau proses pelatihan dan menentukan bobot terbaik, sedangkan data uji hanya digunakan pada tahap akhir untuk mengevaluasi kinerja model secara objektif.

### C. Pre-pemrosesan dan Augmentasi

Pada tahap pra-pemrosesan, seluruh citra ultrasonografi (USG) diseragamkan agar sesuai dengan kebutuhan masukan model. Setiap citra dikonversi ke format RGB dan ukurannya diubah menjadi 224×224 piksel sehingga konsisten dengan spesifikasi input EfficientNet-B0. Selanjutnya, nilai intensitas piksel dinormalisasi agar sebanding dengan distribusi data pada tahap *pretraining*, sehingga proses optimasi selama pelatihan menjadi lebih stabil. Tahap pra-pemrosesan ini diterapkan secara konsisten pada seluruh subset, yaitu data latih, validasi,

dan uji, untuk menjaga keseragaman distribusi data antar tahap eksperimen.

Khusus pada data latih, dilakukan augmentasi berupa pembalikan horizontal acak dan rotasi acak kecil sekitar  $\pm 10^\circ$  untuk meningkatkan keragaman data tanpa mengubah karakteristik anatomi utama hati secara signifikan. Transformasi tersebut dipilih karena masih realistis terhadap variasi orientasi dan posisi akuisisi yang umum terjadi pada pemeriksaan USG, sehingga model didorong untuk mempelajari fitur yang lebih robust terhadap perubahan sudut pandang dan pergeseran posisi probe. Contoh hasil transformasi pada tahap pra-pemrosesan dan augmentasi ditunjukkan pada Gbr. 3.



Gbr. 3 Contoh tahapan pra-pemrosesan dan augmentasi citra USG

#### D. Arsitektur Model

Model yang digunakan dalam penelitian ini adalah EfficientNet-B0, yaitu arsitektur *convolutional neural network* (CNN) yang dikembangkan untuk memperoleh keseimbangan antara efisiensi komputasi dan kemampuan ekstraksi fitur. Arsitektur ini dipilih karena menerapkan strategi *compound scaling* yang menyesuaikan kedalaman, lebar, dan resolusi jaringan secara terintegrasi, sehingga dapat memberikan performa yang baik dengan kompleksitas model yang relatif ringan. Pada penelitian ini, model diawali dengan bobot pralatih (*pretrained weights*) dari dataset citra umum melalui skema *transfer learning*. Masukan model berupa citra RGB berukuran  $224 \times 224$  piksel, sedangkan lapisan klasifikasi akhirnya diubah menjadi lima neuron keluaran sesuai dengan jumlah kelas stadium fibrosis hati, yaitu F0, F1, F2, F3, dan F4. Seluruh implementasi dilakukan menggunakan PyTorch dengan pustaka model siap pakai agar konfigurasi arsitektur dan proses pelatihan dapat diterapkan secara konsisten. Ringkasan konfigurasi utama model ditampilkan pada Tabel II.

TABEL II  
RINGKASAN KONFIGURASI MODEL YANG DIGUNAKAN

| Aspek            | EfficientNet-B0                                        |
|------------------|--------------------------------------------------------|
| Jenis arsitektur | Convolutional Neural Network (CNN)                     |
| Ukuran input     | $3 \times 224 \times 224$ piksel (RGB)                 |
| Bobot awal       | Pralatih ( <i>pretrained</i> ) dari dataset citra umum |

|                     |                              |
|---------------------|------------------------------|
| Jumlah kelas output | 5 kelas (F0, F1, F2, F3, F4) |
|---------------------|------------------------------|

#### E. Skema Pelatihan

Untuk mengatasi ketidakseimbangan distribusi kelas pada data fibrosis, proses pelatihan menggunakan fungsi *loss* CrossEntropyLoss dengan pembobotan kelas. Bobot tersebut dihitung berdasarkan distribusi data latih, sehingga kelas yang memiliki jumlah sampel lebih sedikit memperoleh penalti yang lebih besar selama proses optimasi. Rincian bobot kelas untuk F0 - F4 beserta parameter pelatihan utama disajikan pada Tabel III. Model dilatih menggunakan AdamW dengan *learning rate* awal  $1 \times 10^{-4}$ , batch size 32, dan 20 epoch. Selama pelatihan, digunakan *scheduler* ReduceLROnPlateau untuk menurunkan *learning rate* ketika *validation loss* tidak lagi menunjukkan perbaikan, sehingga proses optimasi dapat berlangsung lebih stabil. Pada setiap epoch, performa model dipantau melalui data validasi, dan bobot terbaik disimpan berdasarkan akurasi validasi tertinggi untuk selanjutnya digunakan pada tahap pengujian dengan data uji. Dengan pengaturan ini, pelatihan tidak hanya ditujukan untuk memperoleh akurasi yang tinggi, tetapi juga untuk menjaga kestabilan pembelajaran pada kondisi distribusi data yang tidak seimbang.

TABEL III  
HYPERPARAMETER TRAINING

| Hyperparameter         | Nilai / Pengaturan                                  |
|------------------------|-----------------------------------------------------|
| Fungsi loss            | CrossEntropyLoss berbobot kelas                     |
| Bobot kelas (F0 - F4)  | [0,5981, 1,4680, 1,5950, 1,4753, 0,7451]            |
| Optimizer              | AdamW                                               |
| Learning rate awal     | $1 \times 10^{-4}$                                  |
| Jumlah epoch           | 20                                                  |
| Ukuran batch           | 32                                                  |
| Scheduler              | ReduceLROnPlateau (monitor <i>validation loss</i> ) |
| Kriteria model terbaik | Akurasi validasi tertinggi                          |

#### F. Evaluasi Kinerja Model

Kinerja model dinilai menggunakan data uji yang sepenuhnya dipisahkan dari data latih dan validasi agar hasil evaluasi dapat menggambarkan kemampuan generalisasi model secara lebih objektif. Karena tugas yang dikerjakan berupa klasifikasi multikelas dengan distribusi kelas yang tidak seimbang, penilaian dilakukan dengan beberapa metrik yang saling

melengkapi. Accuracy digunakan untuk melihat proporsi prediksi yang benar terhadap seluruh sampel, sedangkan macro-F1 digunakan untuk menilai performa rata-rata antar kelas secara lebih seimbang. Selanjutnya, macro-AUROC dengan pendekatan *one-vs-rest* dipakai untuk mengukur kemampuan model dalam membedakan setiap kelas berdasarkan skor probabilitas. Selain itu, confusion matrix digunakan untuk mengamati pola kesalahan prediksi antar kelas, terutama pada kelas-kelas fibrosis yang berdekatan. Keandalan probabilitas prediksi juga dianalisis melalui reliability diagram dan Expected Calibration Error (ECE) untuk menilai apakah tingkat keyakinan model sesuai dengan frekuensi prediksi yang benar. Dengan demikian, evaluasi tidak hanya berfokus pada ketepatan klasifikasi, tetapi juga pada kemampuan diskriminasi dan reliabilitas *confidence* model. Rangkuman metrik evaluasi yang digunakan disajikan pada Tabel IV.

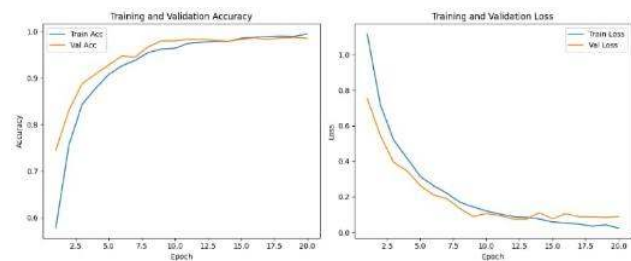
TABEL IV  
METRIK EVALUASI YANG DIGUNAKAN

| Metrik              | Fungsi                                                            |
|---------------------|-------------------------------------------------------------------|
| Accuracy            | Mengukur perbandingan prediksi benar terhadap seluruh sampel      |
| Macro-F1            | Menilai performa rata-rata antar kelas secara seimbang            |
| Macro-AUROC         | Mengukur kemampuan diskriminasi multi-kelas berbasis probabilitas |
| Confusion matrix    | Menganalisis pola kesalahan prediksi antar kelas                  |
| Reliability diagram | Menilai kesesuaian <i>confidence</i> dengan akurasi aktual        |
| ECE                 | Mengukur besar kesalahan kalibrasi probabilitas                   |

### III. HASIL DAN PEMBAHASAN

#### A. Dinamika Pelatihan Model

Dinamika pelatihan model EfficientNet-B0 selama 20 epoch diamati melalui kurva akurasi dan loss pada data latih dan validasi untuk menilai kecepatan konvergensi, kestabilan proses optimasi, serta indikasi overfitting. Sebagaimana ditunjukkan pada Gbr. 4, akurasi data latih dan validasi meningkat secara konsisten sejak epoch awal, sedangkan nilai loss pada kedua subset menurun tajam pada beberapa epoch pertama dan kemudian melandai seiring proses pelatihan. Pola ini menunjukkan bahwa model mampu mempelajari representasi fitur yang relevan dari citra USG dan mencapai kondisi pelatihan yang stabil.



Gbr. 4 Dinamika pelatihan model EfficientNet-B0

Berdasarkan Gbr. 4, akurasi validasi meningkat cepat hingga melampaui 0,90 pada fase awal pelatihan dan kemudian stabil mendekati nilai maksimum pada epoch-epoch akhir. Pada saat yang sama, kurva *loss* train dan validasi terus menurun dengan selisih yang relatif kecil, sehingga tidak terlihat indikasi *overfitting* yang kuat. Secara umum, EfficientNet-B0 menunjukkan konvergensi yang baik dan kestabilan pelatihan yang memadai, sehingga model layak digunakan untuk evaluasi lebih lanjut pada data uji.

#### B. Performa pada Data Uji

Evaluasi pada data uji sebanyak 949 citra menunjukkan bahwa model EfficientNet-B0 mampu mencapai performa klasifikasi yang sangat tinggi, sebagaimana disajikan pada Tabel V. Pada pengujian ini, model memperoleh accuracy 0,9779, macro precision 0,9681, macro recall 0,9664, macro-F1 0,9671, dan macro-AUROC 0,9983. Hasil tersebut menunjukkan bahwa EfficientNet-B0 memiliki kemampuan yang baik dalam mengklasifikasikan lima stadium fibrosis hati pada citra USG, baik dari sisi ketepatan klasifikasi maupun kemampuan diskriminasi berbasis probabilitas.

TABEL V  
PERFORMA KLASIFIKASI MODEL PADA DATA UJI

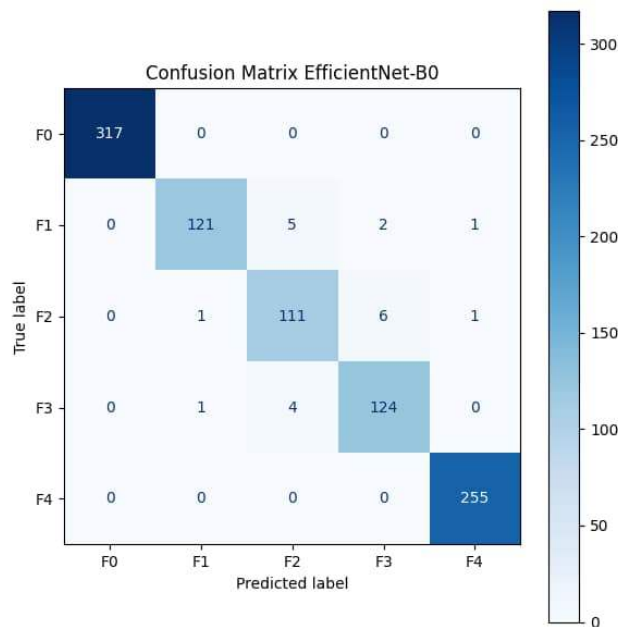
| Metrik (test set)            | EfficientNet-B0 |
|------------------------------|-----------------|
| Accuracy                     | 0,9779          |
| Macro precision              | 0,9681          |
| Macro recall                 | 0,9664          |
| Macro-F1                     | 0,9671          |
| Macro-AUROC                  | 0,9983          |
| Jumlah kesalahan klasifikasi | 21 / 949        |

Berdasarkan Tabel V, EfficientNet-B0 tidak hanya menunjukkan akurasi yang tinggi, tetapi juga performa yang relatif seimbang antar kelas, sebagaimana tercermin dari nilai macro-F1 dan macro recall yang tetap tinggi. Dari total 949 citra uji, model menghasilkan 21 kesalahan klasifikasi, yang

menunjukkan bahwa sebagian besar prediksi telah dilakukan dengan benar. Secara umum, hasil ini menegaskan bahwa EfficientNet-B0 cukup andal untuk tugas klasifikasi stadium fibrosis hati pada citra USG, meskipun analisis lebih rinci per kelas tetap diperlukan untuk mengetahui pola kesalahan yang masih terjadi pada kelas-kelas tertentu.

#### C. Analisis Per Kelas dan Pola Kesalahan

Analisis per kelas dilakukan menggunakan confusion matrix untuk memahami distribusi prediksi benar dan salah pada setiap stadium fibrosis, sebagaimana ditunjukkan pada Gbr. 5. Berdasarkan hasil pengujian, model EfficientNet-B0 menunjukkan performa yang sangat baik pada kelas ekstrem, yaitu F0 dan F4, dengan seluruh sampel pada kedua kelas tersebut terklasifikasi dengan benar. Pada kelas menengah, performa model tetap tinggi, namun kesalahan prediksi masih muncul terutama pada kelas yang berdekatan. Kelas F1 sebagian besar terprediksi dengan benar, meskipun masih terdapat beberapa sampel yang bergeser ke F2, F3, dan F4. Pada kelas F2, kesalahan dominan terjadi ke F3, sedangkan pada kelas F3 masih ditemukan beberapa sampel yang bergeser ke F2. Pola ini menunjukkan bahwa model telah mampu menangkap urutan stadium fibrosis dengan baik, tetapi masih menghadapi kesulitan pada batas antar kelas intermediate yang memiliki karakteristik visual lebih mirip.



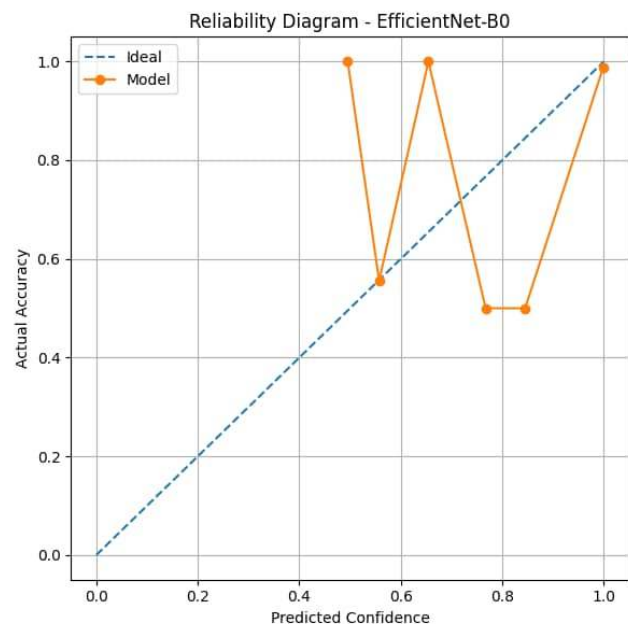
Gbr. 5. Confusion matrix EfficientNet-B0 pada data uji

Berdasarkan Gbr. 5, sebagian besar kesalahan prediksi terjadi antar kelas yang berdekatan dan hampir tidak ditemukan lompatan jauh antar stadium. Hal ini terlihat, misalnya, pada prediksi F2 yang lebih sering bergeser ke F3 dibandingkan langsung ke F4, serta pada F3 yang lebih sering bergeser ke F2

dibandingkan ke kelas lain yang lebih jauh. Pola tersebut menunjukkan bahwa model telah menangkap sifat ordinal stadium fibrosis, meskipun pemisahan antar kelas menengah masih menjadi tantangan. Secara umum, hasil ini memperkuat temuan pada metrik global bahwa EfficientNet-B0 cukup andal untuk klasifikasi stadium fibrosis hati, tetapi peningkatan performa masih diperlukan terutama pada kelas F1 - F3 yang memiliki kemiripan visual lebih tinggi.

#### D. Kalibrasi Probabilitas Model

Selain akurasi klasifikasi, kualitas probabilitas keluaran model juga dievaluasi untuk menilai apakah nilai *confidence* yang dihasilkan benar-benar mencerminkan peluang prediksi yang benar. Analisis ini dilakukan menggunakan reliability diagram dan Expected Calibration Error (ECE) pada data uji, sebagaimana ditunjukkan pada Gbr. 6. Pada pendekatan ini, probabilitas maksimum hasil *softmax* digunakan sebagai *confidence* prediksi, kemudian dikelompokkan ke dalam beberapa rentang untuk dibandingkan antara rata-rata *confidence* dan akurasi aktual. Semakin dekat titik-titik pada *reliability diagram* terhadap garis diagonal ideal dan semakin kecil nilai ECE, semakin baik kualitas kalibrasi model.



Gbr. 6. Reliability diagram model EfficientNet-B0 pada data uji

Berdasarkan Gbr. 6, model EfficientNet-B0 menunjukkan pola kalibrasi yang cukup baik, meskipun masih terdapat deviasi pada beberapa rentang *confidence*. Secara keseluruhan, model menghasilkan nilai ECE sebesar 0,0161, yang menunjukkan bahwa perbedaan antara tingkat keyakinan prediksi dan frekuensi prediksi yang benar relatif kecil. Nilai ini mengindikasikan bahwa probabilitas keluaran model cukup

selaras dengan akurasi aktual, sehingga *confidence* yang dihasilkan dapat dianggap cukup andal. Meskipun demikian, penyimpangan pada beberapa titik dalam *reliability diagram* menunjukkan bahwa kalibrasi model belum sepenuhnya sempurna, sehingga perbaikan lebih lanjut masih dimungkinkan jika model akan digunakan pada skenario keputusan klinis berbasis ambang probabilitas.

#### E. Diskusi dan Implikasi

Hasil penelitian menunjukkan bahwa EfficientNet-B0 mampu menghasilkan performa tinggi dalam klasifikasi stadium fibrosis hati pada citra USG, dengan accuracy 0,9779, macro-F1 0,9671, macro-AUROC 0,9983, dan ECE 0,0161. Model sangat baik dalam mengenali kelas ekstrem F0 dan F4, sedangkan kesalahan masih lebih banyak terjadi pada kelas menengah yang memiliki karakteristik visual lebih mirip. Jika dibandingkan secara kontekstual dengan penelitian terdahulu, hasil ini masih sejalan dengan studi Park et al. yang menunjukkan bahwa EfficientNet memiliki performa kuat pada klasifikasi fibrosis hati berbasis USG. Selain itu, temuan Joo et al. menegaskan bahwa performa model dapat menurun ketika diuji pada perangkat atau domain data yang berbeda, sehingga validasi eksternal tetap diperlukan. Ringkasan perbandingan dengan penelitian sebelumnya disajikan pada Tabel VI.

TABEL VI  
PERBANDINGAN DENGAN PENELITIAN SEBELUMNYA

| Aspek   | Park et al. (2024)    | Joo et al. (2023) | Penelitian ini                      |
|---------|-----------------------|-------------------|-------------------------------------|
| Metode  | EfficientNet, DCNN    | CNN multi-domain  | EfficientNet-B0                     |
| Data    | USG hati              | USG heterogen     | USG hati Kaggle                     |
| Kelas   | F0-F4                 | 5-level           | F0-F4                               |
| Hasil   | AUC 0,96 dan Acc 0,94 | Acc 83-85%        | Acc 0,9779; F1 0,9671; AUROC 0,9983 |
| Catatan | EfficientNet terbaik  | Ada domain bias   | Class weighting dan kalibrasi       |

Berdasarkan Tabel VI, EfficientNet-B0 pada penelitian ini menunjukkan hasil yang kompetitif dibandingkan penelitian sebelumnya. Namun, perbandingan ini bersifat kontekstual karena dataset, skema evaluasi, dan sumber data yang digunakan berbeda. Oleh karena itu, meskipun model menunjukkan potensi sebagai sistem bantu baca USG untuk *staging* fibrosis hati, pengujian eksternal tetap diperlukan sebelum penerapan klinis.

#### IV. KESIMPULAN

Penelitian ini mengevaluasi kinerja EfficientNet-B0 untuk klasifikasi lima stadium fibrosis hati (F0-F4) pada citra ultrasonografi (USG) menggunakan skema *transfer learning* dan penanganan ketidakseimbangan kelas. Hasil pengujian menunjukkan bahwa model mampu mencapai performa yang sangat tinggi, dengan accuracy 0,9779, macro-F1 0,9671, dan macro-AUROC 0,9983 pada data uji. Analisis per kelas menunjukkan bahwa model sangat baik dalam mengenali kelas ekstrem, yaitu F0 dan F4, sedangkan kesalahan prediksi masih lebih sering terjadi pada kelas menengah yang memiliki karakteristik visual lebih mirip. Dari sisi probabilitas prediksi, nilai Expected Calibration Error (ECE) sebesar 0,0161 menunjukkan bahwa model memiliki kalibrasi yang cukup baik sehingga *confidence* yang dihasilkan relatif sejalan dengan akurasi aktual. Secara keseluruhan, hasil ini menunjukkan bahwa EfficientNet-B0 berpotensi digunakan sebagai model inti dalam sistem bantu baca USG untuk *staging* fibrosis hati karena mampu memberikan performa klasifikasi yang tinggi dengan efisiensi komputasi yang baik. Meskipun demikian, penelitian ini masih memiliki keterbatasan berupa penggunaan dataset tunggal dan belum adanya validasi eksternal, sehingga penelitian lanjutan perlu dilakukan pada skenario multi-center untuk menguji kemampuan generalisasi model sebelum diterapkan pada lingkungan klinis.

#### UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada dosen pembimbing, Program Studi Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur, serta semua pihak yang telah memberikan bantuan dan dukungan dalam pelaksanaan penelitian dan penyusunan artikel ini.

#### REFERENSI

- [1] A. Berzigotti et al., “EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis – 2021 update,” *J. Hepatol.*, vol. 75, no. 3, pp. 659–689, Sep. 2021, doi: 10.1016/j.jhep.2021.05.025.
- [2] A. J. Archer, K. J. Belfield, J. G. Orr, F. H. Gordon, and K. W. M. Abeysekera, “EASL clinical practice guidelines: non-invasive liver tests for evaluation of liver disease severity and prognosis,” *Frontline Gastroenterol.*, vol. 13, no. 5, pp. 436–439, Sep. 2022, doi: 10.1136/flgastro-2021-102064.
- [3] Y. N. Zhang et al., “Liver Fibrosis Imaging: A clinical review of Ultrasound and Magnetic Resonance Elastography,” *J. Magn. Reson. Imaging*, vol. 51, no. 1, p. 25, Jan. 2019, doi: 10.1002/jmri.26716.
- [4] C. Fang and P. S. Sidhu, “Ultrasound-based liver elastography: current results and future perspectives,” *Abdom. Radiol. (New York)*, vol. 45, no. 11, p. 3463, Nov. 2020, doi: 10.1007/s00261-020-02717-x.
- [5] S. K. Zhou et al., “A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises,” *Proc. IEEE, Inst. Electr. Electron. Eng.*, vol. 109, no. 5, p. 820, May 2021, doi: 10.1109/JPROC.2021.3054390.
- [6] K. Song, J. Feng, and D. Chen, “A survey on deep learning in medical ultrasound imaging,” *Front. Phys.*, vol. 12, p. 1398393, Jul. 2024, doi: 10.3389/fphy.2024.1398393.
- [7] H. C. Park et al., “Automated classification of liver fibrosis stages using ultrasound imaging,” *BMC Med. Imaging* 2024 241, vol. 24, no. 1, pp. 36–, Feb. 2024, doi: 10.1186/s12880-024-01209-4.

- [8] Y. Joo, H. C. Park, O. J. Lee, C. Yoon, M. H. Choi, and C. Choi, "Classification of Liver Fibrosis From Heterogeneous Ultrasound Image," *IEEE Access*, vol. 11, pp. 9920–9930, 2023, doi: 10.1109/ACCESS.2023.3240216.
- [9] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 24, 2019, *PMLR*. Accessed: Mar. 18, 2026. [Online]. Available: <https://proceedings.mlr.press/v97/tan19a.html>
- [10] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," *Adv. Neural Inf. Process. Syst.*, vol. 32, Oct. 2019, Accessed: Mar. 18, 2026. [Online]. Available: <http://arxiv.org/abs/1902.07208>
- [11] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples".
- [12] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
- [13] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," Jul. 17, 2017, *PMLR*. Accessed: Mar. 18, 2026. [Online]. Available: <https://proceedings.mlr.press/v70/guo17a.html>
- [14] J. Nixon *et al.*, "Measuring Calibration in Deep Learning," Aug. 2020, Accessed: Mar. 18, 2026. [Online]. Available: <http://arxiv.org/abs/1904.01685>