ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Komparasi EfficientNetV2-S dan ConvNeXt-Tiny untuk Klasifikasi Napas Normal-Abnormal pada ICBHI 2017

Ade Rizky Panjaitan[1*], Fetty Tri Anggraeny[2], Eva Yulia Puspaningrum[3]

[1,2,3] Informatika, Universitas Pembangunan Nasional "Veteran"Jawa Timur
[2]fettyanggraeny.if@upnjatim.ac.id
[3]evapuspaningrum.if@upnjatim.ac.id

*Corresponding author email: aderizkypan@gmail.com

Abstrak*— Klasifikasi suara pernapasan berbasis kecerdasan buatan berpotensi mendukung deteksi dini gangguan pernapasan secara lebih objektif dan konsisten. Penelitian ini bertujuan membandingkan kinerja EfficientNetV2-S dan ConvNeXt-Tiny untuk klasifikasi siklus napas pada dataset ICBHI 2017 menggunakan representasi log-mel spectrogram. Data diproses pada tingkat respiratory cycle dan dibagi menggunakan subject-independent split 60:40, dengan 20% dari train pool digunakan sebagai data validasi. Model dilatih pada skema empat kelas, yaitu normal, crackle, wheeze, dan both, kemudian hasil prediksi dipetakan ke skema dua kelas, yaitu normal dan abnormal. Tahap praproses meliputi segmentasi audio, resampling, normalisasi, augmentasi audio, pembentukan log-mel spectrogram, dan augmentasi citra berbasis masking waktu-frekuensi. Hasil evaluasi pada checkpoint best validation sensitivity menunjukkan bahwa pada skema empat kelas EfficientNetV2-S memperoleh specificity 69,16%, accuracy 52,85%, dan AUC OVR Macro 69,13%, sedangkan ConvNeXt-Tiny memperoleh sensitivity 38,04% dan AUC OVR Macro 70,17%. Pada skema dua kelas dengan threshold terbaik dari data validasi, EfficientNetV2-S menghasilkan sensitivity 54,30%, specificity 68,06%, harmonic score 60,41%, dan accuracy 62,04%, sedangkan ConvNeXt-Tiny menghasilkan sensitivity 70,96%, specificity 50,55%, harmonic score 59,04%, dan accuracy 59,48%. Hasil penelitian menunjukkan bahwa kedua model lebih efektif untuk klasifikasi normal-abnormal dibandingkan klasifikasi rinci empat kelas. EfficientNetV2-S cenderung memberikan performa yang lebih seimbang, sedangkan ConvNeXt-Tiny lebih unggul dalam sensitivitas deteksi abnormalitas.

Kata Kunci— **suara pernapasan, ICBHI 2017, log-mel spectrogram, EfficientNetV2-S, ConvNeXt-Tiny, klasifikasi normal-abnormal.**

I. PENDAHULUAN

Penyakit pernapasan masih menjadi persoalan kesehatan penting secara global. Organisasi Kesehatan Dunia melaporkan bahwa *chronic obstructive pulmonary disease* (COPD) merupakan penyebab kematian keempat di dunia dan menyebabkan 3,5 juta kematian pada tahun 2021, atau sekitar 5% dari seluruh kematian global [1]. Data tersebut menunjukkan bahwa gangguan pernapasan masih memberikan beban kesehatan yang besar, sehingga deteksi dini terhadap kelainan suara napas memiliki nilai klinis yang penting. Auskultasi sebagai pemeriksaan awal bersifat non-invasif, cepat, dan relatif murah, tetapi interpretasinya masih bergantung pada pengalaman tenaga medis sehingga hasilnya dapat bersifat subjektif [2] [3]. Perkembangan stetoskop digital, pemrosesan sinyal, dan kecerdasan buatan mendorong pengembangan sistem bantu diagnosis yang lebih objektif untuk membedakan suara napas normal dan abnormal. Dalam penelitian klasifikasi suara pernapasan, dataset ICBHI 2017 menjadi salah satu acuan utama karena memuat 920 rekaman dari 126 partisipan dengan total 6898 respiratory cycle yang dianotasi ke dalam kategori normal, crackle, wheeze, dan kombinasi keduanya [2].

Berbagai penelitian menunjukkan bahwa representasi waktu-frekuensi efektif untuk menonjolkan pola penting pada suara pernapasan. Bardou dkk. [3] menunjukkan bahwa CNN mampu mengungguli metode berbasis fitur buatan tangan, sedangkan Asatani dkk. [4] [5] mengembangkan pendekatan berbasis spectrogram dan CRNN dengan hasil yang baik pada ICBHI 2017. Studi lain juga memperlihatkan bahwa jenis representasi seperti spectrogram, scalogram, melspectrogram, dan gammatonegram memengaruhi performa klasifikasi [6]. Ariyanti dkk. [7] selanjutnya menunjukkan bahwa representasi berbasis log-mel spectrogram juga efektif sebagai masukan model visi modern. Selain itu, pendekatan CNN pralatih dengan skema transfer learning telah banyak digunakan pada berbagai tugas klasifikasi citra karena mampu meningkatkan efisiensi pelatihan dan tetap mempertahankan kemampuan ekstraksi fitur visual yang kuat. Hasan dkk. [8] menunjukkan bahwa pemanfaatan VGG-16 berbasis ImageNet dapat digunakan secara efektif pada tugas klasifikasi citra, sedangkan Idris dkk. [9] menunjukkan bahwa model CNN pre-trained juga efektif pada domain citra medis. Temuan tersebut memperkuat relevansi penggunaan backbone CNN modern dalam penelitian ini untuk mengolah representasi visual berupa log-mel spectrogram.

Dalam konteks ini, ConvNeXt-Tiny dan EfficientNetV2-S dipilih karena memiliki keunggulan yang relevan dengan studi kasus klasifikasi respiratory cycle berbasis citra log-mel spectrogram. ConvNeXt-Tiny merupakan arsitektur ConvNet modern yang dirancang untuk menghasilkan representasi fitur visual yang kuat, sehingga berpotensi lebih baik dalam menangkap pola tekstur halus pada citra log-mel spectrogram, termasuk perbedaan antar suara napas abnormal yang memiliki karakteristik mirip [10]. Sementara itu, EfficientNetV2-S dirancang untuk meningkatkan efisiensi parameter dan kecepatan pelatihan, sehingga berpotensi sesuai untuk proses transfer learning pada dataset yang relatif terbatas [11]. Perbedaan karakteristik tersebut penting karena klasifikasi

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

suara pernapasan tidak hanya menuntut akurasi, tetapi juga sensitivitas terhadap kelas abnormal serta efisiensi komputasi. Oleh karena itu, penelitian ini bertujuan mengomparasikan kinerja EfficientNetV2-S dan ConvNeXt-Tiny pada dataset ICBHI 2017 menggunakan log-mel spectrogram, sehingga dapat memberikan dasar yang lebih terarah dalam pemilihan model dengan backbone CNN modern untuk pengembangan sistem diagnosis suara pernapasan berbasis kecerdasan buatan.

II. LANDASAN TEORI

Bab ini membahas landasan teori yang mendukung penelitian. Teori-teori yang dibahas meliputi konsep dasar suara pernapasan, dataset ICBHI 2017, representasi log-mel spectrogram, arsitektur Convolutional Neural Network (CNN), EfficientNetV2-S, ConvNeXt-Tiny, serta evaluasi model.

A. Suara Pernapasan dan Dataset ICBHI

Suara pernapasan merupakan sinyal akustik yang dihasilkan selama proses inspirasi dan ekspirasi, sehingga dapat digunakan sebagai indikator kondisi saluran pernapasan. Dalam praktik klinis, suara pernapasan dibedakan menjadi suara normal dan abnormal, di mana suara abnormal dapat berupa crackle, wheeze, atau kombinasi keduanya [2]. Analisis otomatis terhadap suara pernapasan penting dilakukan untuk mengurangi subjektivitas auskultasi dan meningkatkan konsistensi penilaian [2] [3]. Dalam penelitian berbasis kecerdasan buatan, analisis umumnya dilakukan pada tingkat *respiratory cycle*, yaitu satu siklus napas lengkap yang telah diberi anotasi batas awal dan akhirnya [2].

Dataset yang digunakan dalam penelitian ini adalah ICBHI 2017 *Respiratory Sound Database*, yang terdiri atas 920 rekaman dari 126 partisipan dengan total sekitar 6898 *respiratory cycle* [2]. Keberagaman subjek, perangkat, dan kondisi perekaman menjadikan dataset ini relevan sebagai tolok ukur dalam pengembangan sistem klasifikasi otomatis. Pada penelitian ini, label anotasi dibentuk ke dalam empat kelas, yaitu normal, crackle, wheeze, dan both. Selain itu, label biner juga dibentuk dengan mengelompokkan crackle, wheeze, dan both ke dalam kelas abnormal, sehingga hasil prediksi dapat dianalisis baik pada skema empat kelas maupun dua kelas [2] [5].

B. Praproses Sinyal Audio

Sinyal suara pernapasan termasuk sinyal non-stasioner, sehingga memerlukan tahap praproses agar karakteristik pentingnya dapat diekstraksi secara lebih baik [4] [5]. Pada penelitian ini, setiap rekaman terlebih dahulu disegmentasi berdasarkan file anotasi *respiratory cycle*. Setelah itu, sinyal audio diproses melalui resampling ke 16 kHz dan normalisasi amplitudo agar data menjadi lebih seragam antar sampel. Pada data pelatihan, notebook juga menerapkan augmentasi audio untuk meningkatkan keragaman data dan membantu mengatasi ketidakseimbangan kelas. Teknik augmentasi yang digunakan meliputi time-stretch, pitch-shift, dan penambahan Gaussian noise. Strategi ini sejalan dengan tujuan peningkatan generalisasi model pada data suara pernapasan yang memiliki variasi temporal dan spektral cukup tinggi [5] [7].

C. Representasi Log-Mel Spectrogram

Representasi waktu-frekuensi merupakan pendekatan yang umum digunakan untuk analisis suara pernapasan karena mampu menggambarkan perubahan energi sinyal pada sumbu waktu dan frekuensi secara bersamaan [4] [5]. Salah satu representasi yang banyak digunakan adalah log-mel spectrogram, yaitu transformasi spektral yang memetakan frekuensi ke skala mel dan menerapkan kompresi logaritmik agar distribusi energi menjadi lebih stabil. Pada penelitian ini, formulasi log-mel spectrogram yang digunakan ditunjukkan pada persamaan (1) hingga persamaan (3) [7] [11].

$$X(m,k) = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-j2\pi kn/N} \quad (1)$$

$$mel(f) = 2595 log_{10}\left(1 + \frac{f}{700}\right) \quad (2)$$

$$S_{logmel} = log(M \cdot |X(m,k)|^2 + \epsilon) \quad (3)$$

Representasi waktu-frekuensi merupakan pendekatan yang umum digunakan untuk analisis suara pernapasan karena mampu menggambarkan perubahan energi sinyal pada sumbu waktu dan frekuensi secara bersamaan [4] [5]. Salah satu representasi yang banyak digunakan adalah log-mel spectrogram, yaitu transformasi spektral yang memetakan frekuensi ke skala mel dan menerapkan kompresi logaritmik agar distribusi energi menjadi lebih stabil. Dalam penelitian ini, proses pembentukan log-mel spectrogram diawali dengan perhitungan representasi spektral sinyal seperti pada Persamaan (1), dilanjutkan dengan pemetaan frekuensi ke skala mel pada Persamaan (2), dan diakhiri dengan transformasi logaritmik untuk memperoleh representasi log-mel spectrogram pada Persamaan (3) [7].

D. Augmentasi Citra Spectrogram dan Penanganan Ketidakseimbangan Kelas

Selain augmentasi pada level audio, penelitian ini juga menerapkan augmentasi pada citra spectrogram melalui masking pada sumbu frekuensi dan waktu yang sejalan dengan prinsip SpecAugment. Pendekatan ini digunakan untuk menambah variasi data latih, mengurangi overfitting, dan meningkatkan kemampuan generalisasi model [12] [13]. Pada implementasi penelitian ini, augmentasi citra hanya diterapkan pada data latih, sedangkan data validasi dan data uji hanya melalui proses resize dan normalisasi agar evaluasi tetap objektif [14].

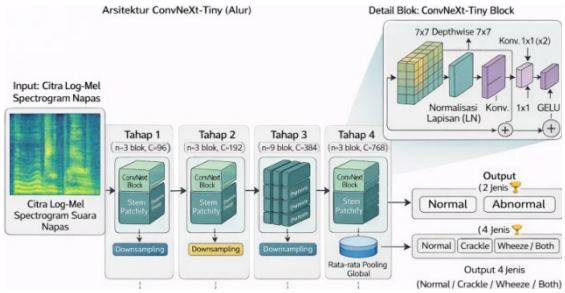
Strategi tersebut juga berkaitan dengan penanganan ketidakseimbangan kelas, yang merupakan salah satu tantangan penting pada dataset ICBHI 2017 karena distribusi jumlah sampel antar kelas tidak merata [2]. Untuk mengatasi kondisi tersebut, penelitian ini menerapkan augmentasi tambahan pada kelas minoritas, penggunaan class-balanced sampler, dan Focal

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Loss. Kombinasi ini diharapkan dapat meningkatkan sensitivitas model terhadap kelas abnormal serta membantu model mempelajari distribusi data secara lebih seimbang [18].

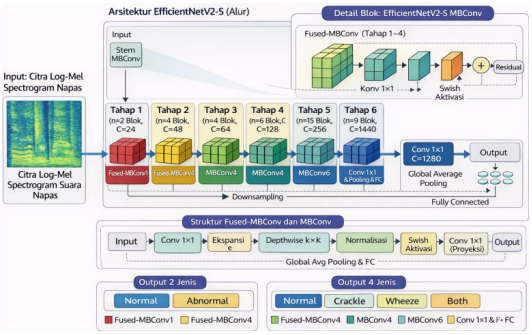
E. Arsitektur CNN untuk Klasifikasi Suara Pernapasan

Convolutional Neural Network (CNN) merupakan arsitektur deep learning yang banyak digunakan dalam klasifikasi citra karena mampu mengekstraksi fitur spasial secara otomatis melalui operasi konvolusi [3]. Dalam klasifikasi suara pernapasan, CNN umumnya bekerja pada representasi visual seperti spectrogram atau log-mel spectrogram, sehingga pola akustik pada domain waktu-frekuensi dapat diperlakukan sebagai pola visual dua dimensi [3] [7]. Penelitian sebelumnya menunjukkan bahwa pendekatan berbasis CNN, spectrogram, dan CRNN efektif untuk klasifikasi suara paru pada dataset ICBHI 2017 [4] [5]. Oleh karena itu, pada penelitian ini digunakan dua arsitektur CNN modern, yaitu ConvNeXt-Tiny dan EfficientNetV2-S, sebagai model pembanding.



Gbr. 1 Arsitektur ConNeXt-Tiny

ConvNeXt-Tiny merupakan arsitektur ConvNet modern yang dikembangkan dengan mengadopsi sejumlah pembaruan desain sehingga tetap kompetitif pada berbagai tugas visi komputer [10]. Arsitektur ini memodernisasi CNN melalui penggunaan depthwise convolution, ukuran kernel yang lebih besar, serta pengaturan blok jaringan yang lebih menyerupai desain model visi modern. Karakteristik tersebut membuat ConvNeXt-Tiny memiliki kemampuan representasi fitur visual yang kuat, sehingga berpotensi lebih baik dalam menangkap pola tekstur halus pada citra log-mel spectrogram, termasuk perbedaan antar suara napas abnormal yang memiliki kemiripan pola frekuensi [10].



Gbr. 2 Arsitektur EfficientNetV2-S

Sementara itu, EfficientNetV2-S merupakan pengembangan dari keluarga EfficientNet yang dirancang untuk meningkatkan efisiensi parameter dan kecepatan pelatihan tanpa mengorbankan performa [11]. Arsitektur ini menggabungkan prinsip penskalaan model yang efisien dengan blok konvolusi yang dioptimalkan untuk mempercepat proses pelatihan dan mendukung transfer learning secara efektif [11]. Dengan karakteristik tersebut menjadikan EfficientNetV2-S relevan untuk penelitian ini, karena dataset ICBHI 2017 memiliki ukuran terbatas dan memerlukan model yang tetap efisien namun mampu mengekstraksi fitur penting dari citra log-mel spectrogram [11].

Pada implementasi penelitian ini, kedua model menggunakan bobot pralatih dari Torchvision, kemudian lapisan klasifikasi akhirnya disesuaikan untuk menghasilkan empat kelas, yaitu normal, crackle, wheeze, dan both [10] [11]. Dengan demikian, perbandingan antara ConvNeXt-Tiny dan EfficientNetV2-S diharapkan dapat menunjukkan perbedaan performa antara arsitektur CNN modern yang berorientasi pada kekuatan representasi fitur dan arsitektur yang berorientasi pada efisiensi pelatihan.

F. Strategi Pelatihan Model

Pelatihan model pada penelitian ini menggunakan AdamW sebagai optimizer, yang merupakan varian Adam dengan decoupled weight decay untuk meningkatkan regularisasi [15]. Selain itu, diterapkan warmup learning rate pada epoch awal dan cosine annealing untuk penurunan laju pembelajaran secara bertahap [16] [17]. Fungsi kerugian yang digunakan adalah Focal Loss, yang dirancang untuk meningkatkan perhatian model pada sampel yang sulit diklasifikasikan serta membantu penanganan ketidakseimbangan kelas [18]. Kombinasi strategi tersebut diterapkan secara konsisten pada EfficientNetV2-S dan ConvNeXt-Tiny agar perbandingan kedua model dilakukan dalam pengaturan pelatihan yang sebanding.

G. Evaluasi Kinerja Model

Evaluasi model dilakukan menggunakan beberapa metrik klasifikasi, seperti accuracy, precision, recall, dan confusion matrix. Metrik-metrik tersebut digunakan untuk menilai performa model baik pada skema klasifikasi empat kelas maupun pada hasil pemetaan ke dua kelas, sehingga evaluasi tidak hanya berfokus pada ketepatan prediksi secara keseluruhan, tetapi juga pada kemampuan model dalam membedakan sampel normal dan abnormal secara lebih tepat.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Selain metrik umum berupa accuracy, precision, recall, dan F1-score pada Persamaan (4) sampai Persamaan (7), penelitian ini juga menggunakan metrik yang umum pada studi berbasis dataset ICBHI 2017, yaitu Sensitivity (Se), Specificity (Sp) untuk evaluasi skema empat kelas sebagaimana ditunjukkan pada Persamaan (8) sampai Persamaan (10), serta Sensitivity biner (Se_{bin}), Specificity biner (Sp_{bin}), dan Harmonic Score (HS) untuk evaluasi skema dua kelas sebagaimana ditunjukkan pada Persamaan (11) sampai Persamaan (13) [5] [6]. Pada skema empat kelas, Sensitivity digunakan untuk mengukur kemampuan model dalam mengenali kelas abnormal, yaitu crackle, wheeze, dan both, sedangkan Specificity digunakan untuk mengukur kemampuan model dalam mengenali kelas normal.

$$Se = \frac{P_c + P_w + P_b}{N_c + N_w + N_b} \quad (8)$$

$$Sp = \frac{P_n}{N_n} \quad (9)$$

$$S = \frac{Se + Sp}{2} \quad (10)$$

$$Se_{bin} = \frac{TP}{TP + FN} \quad (11)$$

$$Sp_{bin} = \frac{TN}{TN + FP} \quad (12)$$

$$HS = \frac{2 \times Se_{bin} \times Sp_{bin}}{Se_{bin} + Sp_{bin}} \quad (13)$$

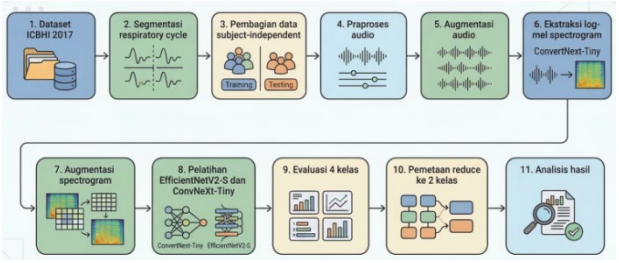
Pada skema dua kelas, Se_{bin} dihitung berdasarkan perbandingan antara true positive (TP) dan jumlah true positive serta false negative (FN), sedangkan SP_{bin} dihitung berdasarkan perbandingan antara true negative (TN) dan jumlah true negative serta false positive (FP). Nilai AS_{bin} diperoleh dari rata-rata Se_{bin} dan Sp_{bin}, sedangkan HS digunakan untuk menunjukkan keseimbangan harmonik antara sensitivitas dan spesifisitas pada klasifikasi biner. Adapun keterangan pada Persamaan (8) sampai Persamaan (10) adalah P_c jumlah prediksi benar pada kelas crackle, P_w jumlah prediksi benar pada kelas wheeze, P_b jumlah prediksi benar pada kelas both, P_n jumlah prediksi benar pada kelas normal, N_c jumlah sampel aktual kelas crackle, N_w jumlah sampel aktual kelas wheeze, N_b jumlah sampel aktual kelas both, dan N_n jumlah sampel aktual kelas normal.

III. METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi penelitian yang digunakan sebagai pedoman eksperimen. Pembahasan meliputi desain penelitian, sumber dan pembagian data, tahap praproses sinyal audio, pembentukan log-mel spectrogram, strategi pelatihan model, serta metode evaluasi kinerja yang digunakan untuk membandingkan EfficientNetV2-S dan ConvNeXt-Tiny.

A. *Desain Penelitian*

Penelitian ini menggunakan pendekatan eksperimen komputasional yaitu untuk membandingkan kinerja model EfficientNetV2-S dan ConvNeXt-Tiny pada klasifikasi suara pernapasan berbasis citra log-mel spectrogram. Model dilatih pada skema empat kelas, yaitu normal, crackle, wheeze, dan both, berdasarkan anotasi *respiratory cycle* pada dataset ICBHI 2017. Selanjutnya, hasil prediksi dipetakan ke dalam skema dua kelas, yaitu normal dan abnormal, untuk mendukung analisis utama penelitian. Alur penelitian ditunjukkan pada Gbr. 3. Pendekatan ini digunakan agar model mempelajari variasi bunyi abnormal secara lebih rinci pada tahap pelatihan, sementara hasil evaluasi tetap dapat dianalisis sesuai tujuan klasifikasi normal-abnormal.



Gbr. 3 Alur Metodologi Penelitian

B. *Dataset dan Pembagian Data*

Dataset yang digunakan dalam penelitian ini adalah ICBHI 2017 *Respiratory Sound Database*, yaitu dataset terbuka yang banyak digunakan pada penelitian klasifikasi suara pernapasan. Dataset ini terdiri atas 920 rekaman dari 126 partisipan dengan total sekitar 6898 *respiratory cycle* yang telah dianotasi ke dalam kategori normal, crackle, wheeze, dan kombinasi crackle serta wheeze [2].

TABEL I
DATASET ICBHI 2017

Kelas	Jumlah Pasien
Normal	3642
Crackle	1864
Wheeze	886
Both (Crackle & Wheeze)	506
Total	6898

Pada penelitian ini, unit analisis yang digunakan adalah *respiratory cycle*, sehingga setiap sampel yang diproses merupakan segmen sinyal hasil pemotongan berdasarkan anotasi waktu awal dan akhir pada setiap siklus napas. Pembagian data dilakukan menggunakan pendekatan subject-independent split agar data dari pasien yang sama tidak muncul pada lebih dari satu subset, sehingga dapat menghindari *data leakage* dan menghasilkan evaluasi yang lebih representatif terhadap kemampuan generalisasi model. Skema pembagian yang digunakan adalah 60:40, yaitu 60% data pasien sebagai train set dan 40% sebagai data uji. Selanjutnya, 20% dari train set dialokasikan sebagai data validasi, sedangkan sisanya digunakan sebagai data latih.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

TABEL II
PEMBAGIAN DATA PENELITIAN

Subset	Jumlah Pasien	Jumlah Segmen
Train	60	3450
Val	15	553
Test	51	2895
Total	126	6898

C. *Segmentasi, Pelabelan, dan Praproses Data*

Tahap awal pengolahan data dilakukan dengan memanfaatkan metadata anotasi untuk memotong rekaman suara pernapasan menjadi segmen-segmen *respiratory cycle*. Setiap segmen diambil berdasarkan informasi waktu mulai dan waktu akhir yang tersedia pada anotasi dataset, kemudian disimpan sebagai sampel individual untuk proses pelatihan dan evaluasi. Pada penelitian ini, pelabelan data dilakukan dalam dua skema, yaitu empat kelas (normal, crackle, wheeze, dan both) yang digunakan pada tahap pelatihan, serta dua kelas (normal dan abnormal) yang dibentuk dengan menggabungkan crackle, wheeze, dan both ke dalam satu kategori abnormal. Setelah segmentasi dilakukan, setiap sinyal audio dipraproses agar memiliki format yang seragam sebelum diekstraksi menjadi fitur, meliputi konversi ke format mono, resampling untuk menyeragamkan frekuensi sampling, serta normalisasi amplitudo untuk mengurangi variasi intensitas antarrekaman.

D. *Augmentasi Data Audio*

Untuk meningkatkan keragaman data latih dan membantu mengatasi ketidakseimbangan kelas, penelitian ini menerapkan augmentasi pada level audio. Teknik augmentasi yang digunakan meliputi perubahan kecepatan sinyal, perubahan tinggi nada, dan penambahan derau acak. Augmentasi hanya diterapkan pada data latih, sedangkan data validasi dan data uji tidak dimodifikasi. Melalui penggunaan augmentasi audio bertujuan untuk menghasilkan variasi sinyal baru yang tetap mempertahankan karakteristik utama suara pernapasan, sehingga model dapat memiliki kemampuan generalisasi yang lebih baik. Selain itu, strategi augmentasi pada data audio dan spectrogram juga telah banyak diterapkan untuk meningkatkan performa model klasifikasi berbasis representasi akustik.

E. *Ekstraksi Fitur Log-Mel Spectrogram*

Setelah melalui tahap praproses, setiap segmen suara pernapasan dikonversi ke dalam representasi log-mel spectrogram. Proses pembentukan log-mel spectrogram dilakukan dengan memperhatikan parameter ukuran jendela analisis, langkah pergeseran waktu, serta jumlah filter pada skala mel. Hasil citra representasi dengan log-mel spectrogram kemudian dikonversi ke dalam bentuk citra tiga kanal dan disesuaikan ukurannya agar kompatibel dengan arsitektur model berbasis visi. Dengan deminkian, melalui proses ini pola akustik pada sinyal suara pernapasan direpresentasikan dalam bentuk citra dua dimensi yang selanjutnya digunakan sebagai masukan model.

F. *Augmentasi Spectrogram dan Penanganan Ketidakseimbangan Kelas*

Selain augmentasi pada level audio, penelitian ini juga menerapkan augmentasi pada citra spectrogram melalui masking pada sumbu frekuensi dan waktu yang sejalan dengan prinsip SpecAugment. Augmentasi ini hanya diterapkan pada data latih setelah proses konversi citra dan normalisasi, sedangkan data validasi dan data uji hanya melalui tahap *resize* dan normalisasi. Di sisi lain, distribusi jumlah sampel pada setiap kelas dalam data suara pernapasan cenderung tidak seimbang, terutama antara kelas normal dan beberapa kelas abnormal tertentu. Sebagaimana ditunjukkan pada Tabel II, jumlah sampel pada kelas minoritas meningkat setelah dilakukan augmentasi pada data latih. Untuk mengatasi ketidakseimbangan tersebut, penelitian ini menerapkan tiga strategi utama, yaitu augmentasi tambahan pada kelas minoritas, penggunaan *class-balanced sampler* agar distribusi sampel dalam setiap batch lebih proporsional, serta *Focal Loss* sebagai fungsi kerugian agar model lebih peka terhadap sampel yang sulit diklasifikasikan [16]. Kombinasi strategi tersebut diharapkan dapat meningkatkan kemampuan model dalam mengenali kelas abnormal secara lebih seimbang.

TABEL III
DISTRIBUSI KELAS DATA LATIH SEBELUM DAN SESUDAH AUGMENTASI

Kelas	Sebelum Augmentasi	Setelah Augmentasi
Normal	1689	1689
Crackle	1130	1689
Wheeze	333	1689
Both	298	1689
Total	3450	6756

G. *Implementasi Model*

Penelitian ini menggunakan dua model klasifikasi citra, yaitu EfficientNetV2-S dan ConvNeXt-Tiny, sebagai model pembanding. Kedua model diimplementasikan menggunakan bobot pralatih (*pretrained weights*) dari Torchvision dan menerima masukan berupa citra log-mel spectrogram hasil praproses. Pada masing-masing model, lapisan klasifikasi akhir disesuaikan agar menghasilkan empat keluaran kelas, yaitu normal, crackle, wheeze, dan both. Prediksi dari skema empat kelas tersebut kemudian dipetakan ke dalam skema dua kelas, yaitu normal dan abnormal, untuk mendukung analisis utama penelitian. Dengan pendekatan ini, model tetap mempelajari variasi bunyi abnormal secara lebih rinci pada tahap pelatihan, sementara hasil evaluasi juga dapat dianalisis pada tingkat klasifikasi normal-abnormal.

H. *Strategi Pelatihan Model*

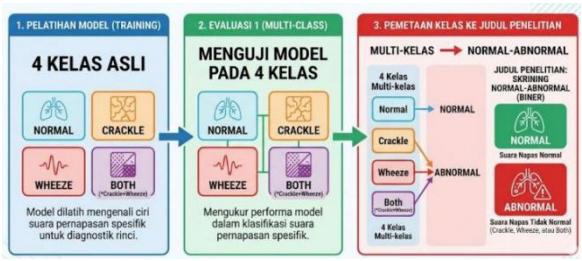
Pelatihan model pada penelitian ini menggunakan AdamW sebagai optimizer. Pada tahap awal pelatihan diterapkan warmup learning rate, kemudian laju pembelajaran diturunkan secara bertahap menggunakan cosine annealing selama proses pelatihan. Fungsi kerugian yang digunakan adalah Focal Loss untuk membantu penanganan ketidakseimbangan kelas. Dengan menggunakan strategi pelatihan yang sama diterapkan

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

pada EfficientNetV2-S dan ConvNeXt-Tiny agar perbandingan performa kedua model dilakukan dalam kondisi eksperimen yang konsisten.

I. *Evaluasi Kinerja Model*

Evaluasi kinerja model pada penelitian ini dilakukan menggunakan beberapa metrik klasifikasi, yaitu accuracy, precision, recall, F1-score, dan confusion matrix. Perhitungan accuracy, precision, recall, dan F1-score ditunjukkan pada Persamaan (4) sampai Persamaan (7). Evaluasi dilakukan pada dua tingkat, yaitu skema empat kelas dan skema dua kelas hasil pemetaan dari prediksi empat kelas. Pada skema empat kelas, model dievaluasi terhadap kelas normal, crackle, wheeze, dan both. Selanjutnya, hasil prediksi tersebut dipetakan ke dalam skema dua kelas, yaitu normal dan abnormal, dengan menggabungkan kelas crackle, wheeze, dan both ke dalam kategori abnormal. Ilustrasi skema reduksi dari empat kelas menjadi dua kelas ditunjukkan pada Gbr. 4.



Gbr. 4 Ilustrasi skema penggabungan 4 kelas ke 2 kelas

Selain metrik umum tersebut, penelitian ini juga menggunakan Sensitivity (*Se*), Specificity (*Sp*) untuk evaluasi pada skema empat kelas sebagaimana ditunjukkan pada Persamaan (8) sampai Persamaan (10). Pada skema ini, Sensitivity digunakan untuk mengukur kemampuan model dalam mengenali kelas abnormal, yaitu crackle, wheeze, dan both, sedangkan Specificity digunakan untuk mengukur kemampuan model dalam mengenali kelas normal. Untuk evaluasi pada skema dua kelas, digunakan Sensitivity biner (Se_{bin}), Specificity biner (Sp_{bin}), Harmonic Score (*HS*) sebagaimana ditunjukkan pada Persamaan (11) sampai Persamaan (13). Pada skema ini, Se_{bin} dihitung berdasarkan perbandingan antara true positive (TP) dan jumlah true positive serta false negative (FN), sedangkan Sp_{bin} dihitung berdasarkan perbandingan antara true negative (TN) dan jumlah true negative serta false positive (FP). Nilai AS_{bin} diperoleh dari rata-rata Se_{bin} dan Sp_{bin}, sedangkan *HS* digunakan untuk menunjukkan keseimbangan harmonik antara sensitivitas dengan spesifisitas pada klasifikasi biner.

IV. HASIL DAN PEMBAHASAN

A. *Hasil Persiapan dan Pembagian Data*

Dataset yang digunakan dalam penelitian ini adalah ICBHI 2017 *Respiratory Sound Database*. Setelah proses segmentasi berdasarkan anotasi *respiratory cycle*, diperoleh 6898 segmen yang berasal dari 126 pasien. Pada skema empat kelas, distribusi data awal terdiri atas 3642 segmen normal, 1864 segmen crackle, 886 segmen wheeze, dan 506 segmen both. Jika dipetakan ke dalam skema dua kelas, jumlah segmen normal adalah 3642, sedangkan segmen abnormal yang merupakan gabungan crackle, wheeze, dan both berjumlah 3256. Pembagian data dilakukan menggunakan pendekatan subject-independent split dengan rasio 60:40, kemudian 20% dari train pool dialokasikan sebagai data validasi. Hasil pembagian menunjukkan bahwa data latih terdiri atas 3450 segmen dari 60 pasien, data validasi terdiri atas 553 segmen dari 15 pasien, dan data uji terdiri atas 2895 segmen dari 51 pasien. Pendekatan ini memastikan bahwa data dari pasien yang sama tidak muncul pada lebih dari satu subset, sehingga evaluasi yang dihasilkan lebih merepresentasikan kemampuan generalisasi model terhadap pasien yang belum pernah dilihat sebelumnya.

Pada data latih sebelum augmentasi, distribusi kelas terdiri atas 1689 segmen normal, 1130 segmen crackle, 333 segmen wheeze, dan 298 segmen both. Untuk mengurangi ketidakseimbangan kelas, augmentasi audio diterapkan pada kelas minoritas sehingga jumlah data latih setelah augmentasi meningkat menjadi 6756 sampel, dengan distribusi 1689 segmen pada setiap kelas, yaitu normal, crackle, wheeze, dan both. Hasil ini menunjukkan bahwa augmentasi berhasil membuat distribusi data latih menjadi lebih seimbang, terutama pada kelas wheeze dan both yang semula memiliki jumlah sampel paling sedikit. Setelah tahap praproses dan ekstraksi fitur, seluruh segmen kemudian dikonversi menjadi citra log-mel spectrogram berukuran 384 × 384 piksel dalam format tiga kanal. Dengan demikian, jumlah citra yang digunakan pada eksperimen adalah 6756 citra untuk data latih, 553 citra untuk data validasi, dan 2895 citra untuk data uji.

B. *Hasil Pelatihan Model*

Model EfficientNetV2-S dan ConvNeXt-Tiny pada penelitian ini sama-sama dibangun menggunakan bobot pralatih dari Torchvision dengan skema partial fine-tuning, optimizer AdamW, scheduler cosine annealing, warmup 5 epoch, dan fungsi kerugian Focal Loss. EfficientNetV2-S memiliki 20.182.612 parameter dengan 19.841.212 parameter trainable, sedangkan ConvNeXt-Tiny memiliki 27.823.204 parameter. Pada EfficientNetV2-S, checkpoint best validation sensitivity diperoleh pada epoch ke-18 dengan validation sensitivity 0,4912, validation specificity 0,8154, harmonic score 0,6131, dan abnormal macro recall 0,4323. Sementara itu, pada ConvNeXt-Tiny, checkpoint best validation sensitivity diperoleh pada epoch ke-12 dengan validation sensitivity 0,5307, validation specificity 0,7569, harmonic score 0,6239, dan abnormal macro recall 0,4348.

Secara umum, hasil pelatihan menunjukkan bahwa kedua model mampu mempelajari pola suara pernapasan abnormal, tetapi memiliki karakteristik konvergensi yang berbeda ketika ditinjau dari checkpoint best validation sensitivity. ConvNeXt-Tiny mencapai sensitivitas validasi yang lebih tinggi dan lebih cepat, tetapi dengan spesifisitas yang lebih rendah. Sebaliknya,

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

EfficientNetV2-S menghasilkan sensitivitas yang sedikit lebih rendah, namun memberikan keseimbangan yang lebih baik antara sensitivitas dan spesifisitas pada proses pelatihan.

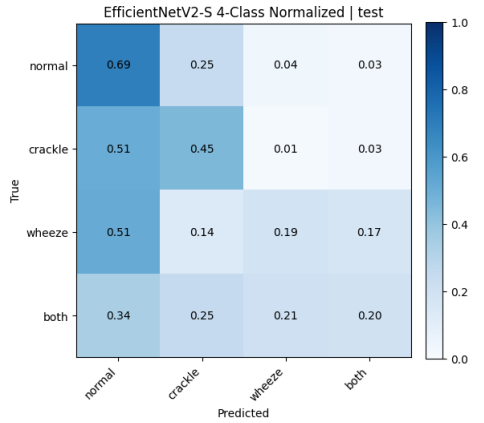
C. *Hasil Evaluasi pada Skema 4 Kelas*

Hasil evaluasi kedua model pada data uji untuk skema empat kelas disajikan pada Tabel III. Pada checkpoint best validation sensitivity, EfficientNetV2-S memperoleh Sensitivity (Se) 31,89%, Specificity (Sp) 69,16%, Harmonic Score (HS) 43,65%, accuracy 52,85%, F1-score Macro 38,24%, dan AUC OVR Macro 69,13%. Sementara itu, ConvNeXt-Tiny memperoleh Sensitivity (Se) 38,04%, Specificity (Sp) 54,98%, Harmonic Score (HS) 44,97%, accuracy 47,56%, F1-score Macro 35,38%, dan AUC OVR Macro 70,17%. Hasil ini menunjukkan bahwa pada skema empat kelas, EfficientNetV2-S lebih baik dalam menjaga pengenalan kelas normal, yang tercermin dari nilai specificity, dan accuracy yang lebih tinggi. Sebaliknya, ConvNeXt-Tiny menunjukkan sensitivity yang lebih tinggi serta AUC OVR Macro yang sedikit lebih baik, sehingga model ini relatif lebih peka dalam mengenali sampel dari kelas abnormal

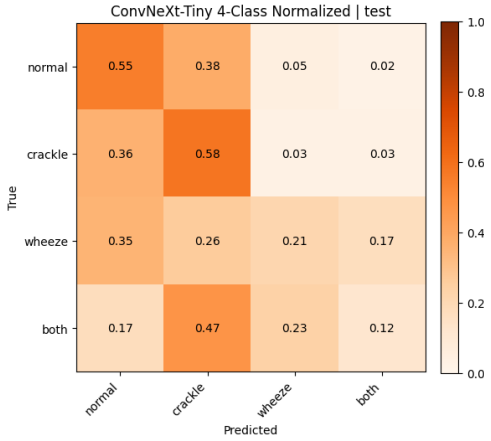
TABEL IV
HASIL EVALUASI DUA MODEL SKEMA 4 KELAS

Model	EfficientNetV2-S	ConvNeXt-Tiny
Sensitivity (Se)	31,89	38,04
Specificity (Sp)	69,16	54,98
Harmonic Score (HS)	43,65	44,97
Accuracy	52,85	47,56
Precision Macro	41,53	38,49
Recall Macro	38,27	36,57
F1-score Macro	38,24	35,38
AUC OVR Macro	69,13	70,17

Jika ditinjau berdasarkan recall per kelas, EfficientNetV2-S memperoleh 69,16% pada kelas normal, 45,32% pada crackle, 18,71% pada wheeze, dan 19,88% pada both. Sementara itu, ConvNeXt-Tiny memperoleh 54,98% pada kelas normal, 58,06% pada crackle, 21,21% pada wheeze, dan 12,05% pada both. Temuan ini menunjukkan bahwa kedua model masih menghadapi kesulitan pada kelas wheeze dan both, yang mengindikasikan bahwa pemisahan antar jenis suara napas abnormal masih menjadi tantangan utama pada skema empat kelas. EfficientNetV2-S cenderung lebih kuat pada kelas normal dan both, sedangkan ConvNeXt-Tiny lebih baik pada kelas crackle dan wheeze. Dengan demikian, perbedaan performa kedua model pada skema empat kelas terutama terletak pada cara masing-masing model menyeimbangkan pengenalan kelas normal terhadap kelas-kelas abnormal yang memiliki karakteristik akustik saling berdekatan. Pola kesalahan klasifikasi kedua model dapat diamati pada confusion matrix di Gbr. 5 dan Gbr. 6, yang menunjukkan bahwa kesalahan prediksi masih didominasi oleh pertukaran antar kelas abnormal.



Gbr. 5 confusion matrix EfficientNetV2-S 4 kelas



Gbr. 6 confusion matrix ConvNeXt-Tiny 4 kelas

D. *Hasil Evaluasi pada Skema 2 Kelas*

Pada tahap selanjutnya, keluaran prediksi pada skema empat kelas dipetakan ke dalam skema dua kelas, yaitu normal dan abnormal, dengan kelas abnormal mencakup crackle, wheeze, dan both. Dalam konteks ini, Sensitivity (Se) didefinisikan sebagai tingkat keberhasilan model dalam mengidentifikasi sampel abnormal, sedangkan Specificity (Sp) didefinisikan sebagai tingkat keberhasilan model dalam mengenali sampel normal. Evaluasi dilakukan menggunakan checkpoint best validation sensitivity pada masing-masing model, dengan nilai threshold yang ditentukan berdasarkan data validasi.

Pada EfficientNetV2-S, threshold terbaik yang diperoleh adalah 0,51, dengan nilai Sensitivity 54,30%, Specificity 68,06%, Harmonic Score 60,41%, accuracy 62,04%, F1-score Macro 61,22%, dan AUC Binary 64,51%. Sementara itu, pada ConvNeXt-Tiny, threshold terbaik juga diperoleh pada 0,51, dengan nilai Sensitivity 70,96%, Specificity 50,55%, Harmonic Score 59,04%, accuracy 59,48%, F1-score Macro 59,45%, dan AUC Binary 66,34%. Hasil tersebut menunjukkan bahwa ConvNeXt-Tiny memiliki kemampuan yang lebih tinggi dalam mendeteksi abnormalitas, sedangkan EfficientNetV2-S

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

menunjukkan performa yang lebih seimbang dalam membedakan kelas normal dan abnormal.

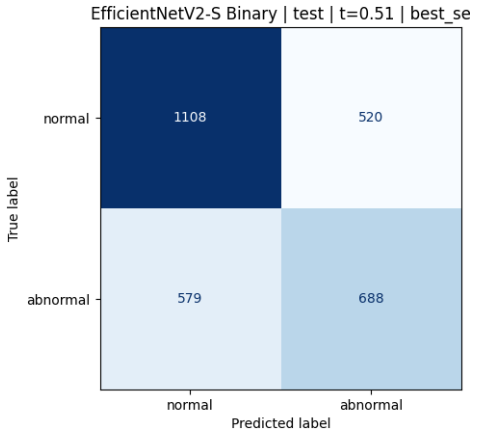
TABEL V
HASIL EVALUASI DUA MODEL SKEMA 2 KELAS

Model	EfficientNetV2-S	ConvNeXt-Tiny
Threshold	0.51	0.51
Sensitivity (Se)	54,30	70,96
Specificity (Sp)	68,06	50,55
Harmonic Score (HS)	60,41	59,04
Accuracy	62,04	59,48
F1 Macro	61,22	59,45
AUC Binary	64,51	66,34

Secara umum, hasil evaluasi pada skema dua kelas menunjukkan performa yang lebih stabil dibandingkan dengan skema empat kelas. Temuan ini mengindikasikan bahwa kedua model lebih mampu membedakan keberadaan abnormalitas suara napas secara umum daripada mengklasifikasikan jenis suara napas abnormal secara lebih rinci. Rincian metrik evaluasi pada skema dua kelas disajikan pada Tabel IV, sedangkan distribusi prediksi benar dan salah dapat diamati melalui confusion matrix pada Gbr. 7 dan Gbr. 8. Hasil ini menegaskan bahwa penyederhanaan klasifikasi dari empat kelas menjadi dua kelas mampu meningkatkan kestabilan performa model. Meskipun demikian, karakteristik kedua model tetap menunjukkan perbedaan, yaitu EfficientNetV2-S yang cenderung lebih seimbang dalam membedakan kelas normal dan abnormal, serta model ConvNeXt-Tiny yang cenderung lebih agresif dalam mendeteksi abnormalitas.

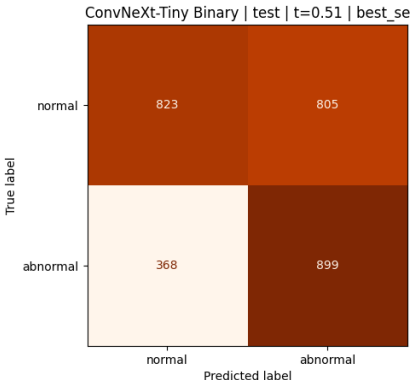
E. *Pembahasan*

Berdasarkan hasil evaluasi pada skema empat kelas, kedua model menunjukkan karakteristik performa yang berbeda. EfficientNetV2-S memperoleh Specificity dan accuracy yang lebih tinggi, sehingga menunjukkan kemampuan yang lebih baik dalam mempertahankan pengenalan kelas normal. Sebaliknya, ConvNeXt-Tiny memperoleh Sensitivity yang lebih tinggi serta AUC OVR Macro yang sedikit lebih baik, sehingga menunjukkan kecenderungan yang lebih kuat dalam menangkap sampel dari kelas abnormal. Temuan ini menunjukkan bahwa perbedaan utama kedua model pada skema empat kelas bukan terletak pada dominasi mutlak salah satu model, melainkan pada perbedaan karakter dalam menyeimbangkan pengenalan kelas normal dan kelas abnormal. Meskipun demikian, kedua model masih menunjukkan keterbatasan pada kelas wheeze dan both. Hal ini mengindikasikan bahwa pemisahan antar jenis suara napas abnormal masih menjadi tantangan utama, meskipun telah diterapkan augmentasi data dan penanganan ketidakseimbangan kelas. Dengan kata lain, representasi log-mel spectrogram telah membantu model dalam mempelajari pola akustik suara napas, tetapi belum sepenuhnya mampu mengatasi kemiripan karakteristik antar kelas abnormal. Kondisi tersebut tercermin dari confusion matrix pada Gbr. 5 dan Gbr. 7, yang menunjukkan bahwa kesalahan prediksi masih didominasi oleh pertukaran antar kelas abnormal.



Gbr. 7 confusin matrix EfficientNetV2-S 2 kelas

Pada evaluasi skema dua kelas, performa kedua model cenderung lebih stabil dibandingkan skema empat kelas. Hasil ini menunjukkan bahwa EfficientNetV2-S maupun ConvNeXt-Tiny lebih efektif dalam mendeteksi keberadaan abnormalitas suara napas secara umum daripada membedakan jenis abnormalitas secara spesifik. Dengan menggunakan checkpoint Best_SE dan threshold terbaik dari data validasi, EfficientNetV2-S menghasilkan Sensitivity 54,30% dan Specificity 68,06%, sedangkan ConvNeXt-Tiny menghasilkan Sensitivity 70,96% dan Specificity 50,55%. Temuan ini menunjukkan bahwa ConvNeXt-Tiny memiliki kemampuan deteksi abnormalitas yang lebih tinggi, tetapi diikuti oleh penurunan kemampuan dalam mengenali kelas normal. Sebaliknya, EfficientNetV2-S memberikan performa yang lebih seimbang dalam membedakan kelas normal dan abnormal.



Gbr. 8 confusion matrix ConvNeXt-Tiny 2 kelas

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kedua model memiliki keunggulan yang berbeda. Pada skema empat kelas, EfficientNetV2-S lebih baik dalam menjaga pengenalan kelas normal, sedangkan ConvNeXt-Tiny lebih sensitif terhadap abnormalitas. Pada skema dua kelas, kedua model menunjukkan performa yang lebih stabil, tetapi tetap

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

mempertahankan karakter yang sama, yaitu EfficientNetV2-S lebih seimbang dalam membedakan kelas normal dan abnormal, sedangkan ConvNeXt-Tiny lebih agresif dalam mendeteksi abnormalitas. Oleh karena itu, pemilihan model perlu disesuaikan dengan tujuan sistem, apakah lebih menekankan keseimbangan klasifikasi atau sensitivitas deteksi abnormalitas pada tahap skrining awal.

F. *Keterbatasan Hasil*

Penelitian ini masih memiliki beberapa keterbatasan yang perlu diperhatikan dalam menafsirkan hasil. Pertama, evaluasi hanya dilakukan pada satu skema pembagian data, yaitu subject-independent split 60:40 dengan 20% dari train pool digunakan sebagai data validasi, sehingga kestabilan performa model pada skenario pembagian data lain belum dapat dipastikan. Kedua, model dilatih pada skema empat kelas, sedangkan evaluasi juga dianalisis melalui pemetaan ke skema dua kelas. Kondisi ini menyebabkan interpretasi hasil perlu mempertimbangkan dua tingkat kompleksitas klasifikasi, yaitu kemampuan model dalam membedakan jenis abnormalitas secara rinci dan kemampuannya dalam mendeteksi abnormalitas secara umum. Ketiga, hasil evaluasi menunjukkan bahwa performa klasifikasi pada skema empat kelas masih terbatas, terutama pada kelas wheeze dan both. Hal ini mengindikasikan bahwa pemisahan antar kelas abnormal masih menjadi tantangan utama, meskipun telah diterapkan augmentasi data dan penanganan ketidakseimbangan kelas. Keempat, pemilihan checkpoint utama yang didasarkan pada best validation sensitivity membuat konfigurasi eksperimen lebih menekankan kemampuan deteksi abnormalitas, sehingga karakter performa model cenderung sensitif terhadap trade-off antara sensitivity dan specificity. Selain itu, penelitian ini masih terbatas pada dataset ICBHI 2017, sehingga kemampuan generalisasi model pada dataset eksternal atau kondisi perekaman yang berbeda belum dapat dipastikan. Oleh karena itu, penelitian lanjutan perlu dilakukan dengan skema validasi yang lebih beragam, pengujian pada data eksternal, serta eksplorasi pendekatan yang mampu meningkatkan pemisahan antar kelas abnormal, khususnya pada kelas wheeze dan both.

V. KESIMPULAN

Penelitian ini menunjukkan bahwa EfficientNetV2-S dan ConvNeXt-Tiny memiliki karakteristik performa yang berbeda dalam klasifikasi siklus napas berbasis log-mel spectrogram pada dataset ICBHI 2017. Pada skema empat kelas, EfficientNetV2-S lebih baik dalam menjaga pengenalan kelas normal, yang ditunjukkan oleh specificity 69,16% dan accuracy 52,85%, sedangkan ConvNeXt-Tiny lebih baik dalam mendeteksi abnormalitas, dengan sensitivity 38,04% dan AUC OVR Macro 70,17%. Pada skema dua kelas, kedua model menunjukkan performa yang lebih stabil, dengan EfficientNetV2-S memberikan hasil yang lebih seimbang melalui specificity 68,06%, harmonic score 60,41%, dan accuracy 62,04%, sementara ConvNeXt-Tiny menunjukkan sensitivity 70,96% yang lebih tinggi untuk deteksi abnormalitas. Secara keseluruhan, kedua model lebih efektif untuk klasifikasi normal-abnormal dibandingkan klasifikasi rinci empat kelas, dengan EfficientNetV2-S lebih sesuai untuk sistem yang menuntut keseimbangan klasifikasi, sedangkan ConvNeXt-Tiny lebih sesuai untuk sistem yang memprioritaskan sensitivitas deteksi abnormalitas.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah berkontribusi dalam penyelesaian penelitian ini. Apresiasi khusus diberikan kepada penyelenggara SANTIKA atas ketersediaan format dan petunjuk penulisan yang sangat membantu dalam penyusunan naskah. Penulis juga berterima kasih kepada berbagai pihak yang telah memberikan dorongan, perhatian, dan saran konstruktif, baik secara langsung maupun melalui dukungan tidak langsung, selama seluruh rangkaian penelitian berlangsung. Semoga karya ini dapat memberi nilai guna dan menjadi rujukan yang bermanfaat bagi pembaca maupun peneliti selanjutnya.

REFERENSI

[1] World Health Organization, “Chronic obstructive pulmonary disease (COPD),” Nov. 6, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd). [Accessed: day-month-year]

[2] B. M. Rocha et al., “An open access database for the evaluation of respiratory sound classification algorithms,” *Physiol. Meas*., vol. 40, no. 3, p. 035001, 2019, doi: 10.1088/1361-6579/ab03ea.

[3] D. Bardou, K. Zhang, and S. M. Ahmad, “Lung sounds classification using convolutional neural networks,” *Artif. Intell. Med*., vol. 88, pp. 58-69, 2018, doi: 10.1016/j.artmed.2018.04.008.

[4] N. Asatani, T. Kamiya, S. Mabu, and S. Kido, “Classification of Respiratory Sounds Using Two Resolution Spectrograms and TF-CRNN,” in *Proc. 33rd Annu. Meeting Biomed. Fuzzy Syst. Assoc*., 2020, pp. 64-67. [Online]. Available: https://www.jstage.jst.go.jp/article/pacbfsa/33/0/33_64/_pdf/-char/en

[5] N. Asatani, T. Kamiya, S. Mabu, and S. Kido, “Classification of respiratory sounds using improved convolutional recurrent neural network,” *Comput. Electr. Eng.*, vol. 94, art. no. 107367, 2021, doi: 10.1016/j.compeleceng.2021.107367.

[6] Z. Neili and K. Sundaraj, “A comparative study of the spectrogram, scalogram, melspectrogram and gammatonegram time-frequency representations for the classification of lung sounds using the ICBHI database based on CNNs,” *Biomed. Eng. / Biomed. Tech*., vol. 67, no. 5, pp. 367-390, 2022, doi: 10.1515/bmt-2022-0180.

[7] W. Ariyanti, K.-C. Liu, K.-Y. Chen, and Y. Tsao, “Abnormal Respiratory Sound Identification Using Audio-Spectrogram Vision Transformer,” *arXiv preprint arXiv:2405.08342*, 2024, doi: 10.48550/arXiv.2405.08342.

[8] F. Hasan, E. B. Priambudi, M. R. Rahmanda, and E. Y. Puspaningrum, “Aplikasi Android untuk Klasifikasi Motif Batik Nitik Yogyakarta Menggunakan VGG-16 dan ImageNet,” *J. Teknol. Inf. dan Komun*. (SCAN), vol. 20, no. 1, pp. 32-41, 2025, doi: 10.33005/scan.v20i1.5644.

[9] M. Idris, F. T. Anggraeny, and R. Mumpuni, “Identifikasi Kanker Paru-Paru pada Gambar Histopatologi Menggunakan Metode Convolutional Neural Network,” *J. Teknol. Inf. dan Komun. (SCAN),* vol. 18, no. 3, pp. 37-44, 2023, doi: 10.33005/scan.v18i3.4728.

[10] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 11966-11976, doi: 10.1109/CVPR52688.2022.01167.

[11] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *arXiv preprint arXiv:2104.00298*, 2021, doi: 10.48550/arXiv.2104.00298.

[12] K. M. Lim, C. P. Lee, Z. Y. Lee, and A. Alqahtani, "EnViTSA: Ensemble of Vision Transformer with SpecAugment for Acoustic Event Classification," *Sensors*, vol. 23, no. 22, p. 9084, 2023, doi: 10.3390/s23229084.

[13] S. Hamdi, M. Oussalah, A. Moussaoui, and M. Saidi, "Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound," *J. Intell. Inf. Syst.*, vol. 59, pp. 367-389, 2022, doi: 10.1007/s10844-022-00707-7.

[14] T. Truong, M. Lenga, A. Serrurier, and S. Mohammadi, "Fused Audio Instance and Representation for Respiratory Disease Detection," *Sensors*, vol. 24, no. 19, p. 6176, 2024, doi: 10.3390/s24196176.

[15] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *arXiv preprint arXiv:1711.05101*, 2019, doi: 10.48550/arXiv.1711.05101.

[16] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour," *arXiv preprint arXiv:1706.02677,* 2017, doi: 10.48550/arXiv.1706.02677.

[17] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," *arXiv preprint arXiv:1608.03983*, 2016, doi: 10.48550/arXiv.1608.03983.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection*," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV),* 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.324.