

# Klasifikasi Suara Paru-Paru Menggunakan Mel Frequency Cepstral Coefficient dan Convolutional Neural Network

## *Lung Sound Classification Using Mel Frequency Cepstral Coefficients and Convolutional Neural Networks*

Sayyidis Syariful Halim, Bulkis Kanata, Lalu A. Syamsul Irfan Akbar

Universitas Mataram, Mataram, Indonesia

### Informasi Artikel:

Diterima: 22 September 2024, Direvisi: 3 November 2024, Disetujui: 19 November 2024

---

#### Abstrak-

**Latar Belakang:** Tantangan dalam diagnosis gangguan pernapasan sering disebabkan oleh kurangnya alat teknologi yang mampu mengenali pola suara paru-paru secara akurat, sehingga dapat mengurangi potensi kesalahan diagnosis subjektif oleh tenaga medis.

**Tujuan:** Penelitian ini bertujuan mengembangkan model klasifikasi suara paru-paru yang dapat mendeteksi gangguan pernapasan secara dini dan akurat.

**Metode:** Metode yang digunakan adalah kombinasi teknik augmentasi data dan ekstraksi fitur Mel Frequency Cepstral Coefficient (MFCC) untuk meningkatkan kinerja Convolutional Neural Network (CNN) dalam mengklasifikasikan suara paru-paru. Sebanyak 1.350 rekaman audio paru-paru dikategorikan ke dalam sembilan kelas, termasuk suara normal dan abnormal. Untuk memperkaya data pelatihan, diterapkan berbagai teknik augmentasi, seperti penambahan *white noise*, *pitch scaling*, *time stretching*, dan *random gain*.

**Hasil:** Hasil penelitian menunjukkan bahwa model E-CNN2D mencapai akurasi hingga 95%, melampaui model sebelumnya dengan rentang akurasi 83-93%.

**Kesimpulan:** Penelitian ini memiliki potensi sebagai solusi alat bantu diagnosis yang cepat dan akurat, sehingga dapat mendukung tenaga medis dalam mengurangi risiko kesalahan diagnosis subjektif pada gangguan pernapasan.

**Kata Kunci:** *Convolution Neural Network*, Deteksi Gangguan Pernapasan, Klasifikasi Suara Paru, *Mel Frequency Cepstral Coefficient*.

---

#### Abstract-

**Background:** Challenges in diagnosing respiratory disorders are often caused by the lack of technological tools capable of accurately recognizing lung sound patterns, which would reduce the potential for subjective misdiagnosis by medical personnel.

**Objective:** This study aims to develop a lung sound classification model that can accurately and early detect respiratory disorders.

**Methods:** The method combines data augmentation techniques and Mel Frequency Cepstral Coefficient (MFCC) feature extraction to improve the Convolutional Neural Network (CNN) performance in classifying lung sounds. A total of 1,350 lung audio recordings were categorized into nine classes, including normal and abnormal sounds. The augmentation techniques applied include the addition of white noise, pitch scaling, time stretching, and random gain to enrich the variety of training data.

**Result:** The results show that the E-CNN2D model can achieve an accuracy of up to 95%, surpassing the previous model, which had an 83-93% accuracy range.

**Conclusion:** These results suggest that this study has the potential to be a fast and accurate diagnostic tool that can help medical personnel reduce the risk of subjective misdiagnosis in respiratory disorders.

**Keywords:** *Convolution Neural Network*, Lung Sound Classification, Mel Frequency Cepstral, Respiratory Disorder Detection.

---

---

**How to Cite:** S. S. Halim, B. Kanata, and S. I. Akbar, "The Lung Sound Classification Using Mel Frequency Cepstral Coefficients and Convolutional Neural Networks", *Jurnal Bumigora Information Technology (BITe)*, vol. 6, no. 2, hlm. 163-176, Dec. 2024. doi: 10.30812/bite.v6i2.4487

This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

**Penulis Korespondensi:**

Bulkis Kanata,  
Program Studi Teknik Elektro, Universitas Mataram,  
Email: [uqikanata@unram.ac.id](mailto:uqikanata@unram.ac.id)

---

**1. PENDAHULUAN**

Auskultasi merupakan metode yang sering digunakan tenaga medis untuk mendeteksi masalah pada sistem pernapasan pasien dengan menyimak bunyi napas menggunakan stetoskop [1], [2]. Stetoskop memiliki peran penting dalam diagnosis non-invasif penyakit pernapasan [3], [4]. Stetoskop berbiaya rendah, cepat, mudah, portable, dan sekarang dapat merekam suara paru-paru pasien, sehingga suara tersebut dapat dianalisis lebih lanjut [5]. Suara paru-paru normal terdiri atas bronkial, bronkovesikular, vesikular, dan trakeal sedangkan suara abnormal meliputi mengi, stridor, ronki, krekels dan *pleural friction rub*. Deteksi suara paru-paru normal atau abnormal ini sangat membantu dalam diagnosis penyakit paru-paru [6]. Suara trakeal dan bronkial dihasilkan oleh aliran udara turbulen melalui trakea dan bronkus batang utama, terdengar keras dan dapat didengar baik selama inspirasi maupun ekspirasi. Suara vesikular lebih lembut dan terdengar samar, paling jelas terdengar di area dinding dada lainnya, terutama selama inspirasi dan menghilang di awal ekspirasi. Suara bronkovesikular terdengar di antara bronkus batang utama dan saluran udara yang lebih kecil, dengan nada dan durasi yang berada di tengah-tengah antara suara trakeal/bronkial dan suara vesikular di dinding dada lainnya [5], [7]. Mengi adalah suara mendesing yang terjadi saat saluran napas kecil menyempit, umum pada asma. Stridor merupakan suara kasar saat inspirasi, sering terjadi karena obstruksi di saluran napas atas. Ronki adalah suara berdengung terkait penumpukan lendir di saluran napas besar, yang bisa hilang setelah batuk. Krekel terdengar seperti berderak selama inspirasi, disebabkan oleh cairan di alveoli, umum pada pneumonia. *Pleural friction rub* adalah suara menggesek yang timbul akibat peradangan pleura [5], [7].

Meskipun auskultasi memiliki banyak keuntungan, kemampuan untuk menganalisis suara pernapasan di antara tenaga medis bervariasi tergantung pada pengalaman klinis individu, karena hasil pemeriksaan semacam ini bersifat subjektif, sangat bergantung pada pengalaman dan kemampuan pendengaran tenaga medis, kepekaan dan kesamaan pola suara karenanya rentan terhadap kesalahan diagnosis [8], [9]. Meta-analisis [6] menunjukkan bahwa auskultasi paru memiliki sensitivitas rendah (37%) meskipun spesifisitas lebih tinggi (89%) dalam mendeteksi patologi paru. Keterbatasan dalam konsistensi dan objektivitas hasil pemeriksaan ini mendorong perlunya pengembangan alat bantu berbasis teknologi yang mampu mengidentifikasi pola suara paru-paru dengan lebih akurat dan konsisten, terlepas dari subjektivitas pendengaran manusia [10].

Pada penelitian terdahulu, [11] mengklasifikasikan gangguan paru-paru menggunakan metode Convolutional Neural Network (CNN) dengan data spektrogram yang ditransformasikan ke format gambar menggunakan *Viridis Color Map* dan menghasilkan akurasi sebesar 71,15%. Sedangkan penelitian [10] yang menggunakan metode kombinasi CNN, *Recurrent Neural Network (RNN)*, dan *Fully Connected Neural Network* memperoleh data pernapasan dari ICBHI 2017 dan mencapai akurasi yang bervariasi berdasarkan metode dan jenis suara yang diuji.

Penelitian ini mengklasifikasikan jenis suara paru-paru sebanyak sembilan kelas. Data terbagi menjadi dua kelompok utama, yaitu suara paru-paru normal (bronkial, bronkovesikular, vesikular dan trakeal) dan suara paru-paru abnormal (krekels, *Pleural Friction Rub*, ronki, stridor). Suara *Bronkial* merupakan suara napas keras dan kasar yang terdengar pada trakea dan laring selama inspirasi dan ekspirasi. *Bronkovesikular* adalah kombinasi suara dari bronkus dan alveoli, terdengar saat udara mengalir di saluran pernapasan. *Vesicular* merupakan suara lembut dengan frekuensi rendah, terdengar di seluruh permukaan paru-paru. Sementara itu, *Trakeal* adalah suara yang sangat keras dan memiliki frekuensi tinggi, dengan durasi inspirasi dan ekspirasi yang hampir sama [7]. Suara abnormal dimulai dengan *Krekels*, suara letupan pendek yang sering dikaitkan dengan

kondisi paru-paru seperti infeksi atau edema. *Pleural Friction Rub* adalah suara gesekan kering antara dua lapisan pleura, yang muncul saat bernapas dan sering dikaitkan dengan peradangan. *Ronki* terdengar seperti gemuruh akibat adanya produksi lendir berlebih, umumnya disebabkan oleh kondisi seperti asma atau bronkitis. *Stridor* disebabkan oleh obstruksi saluran udara bagian atas, biasanya terdengar pada fase inspirasi. Terakhir, *Mengi* adalah suara bernada tinggi akibat penyempitan saluran udara, umumnya terdengar saat fase ekspirasi [5].

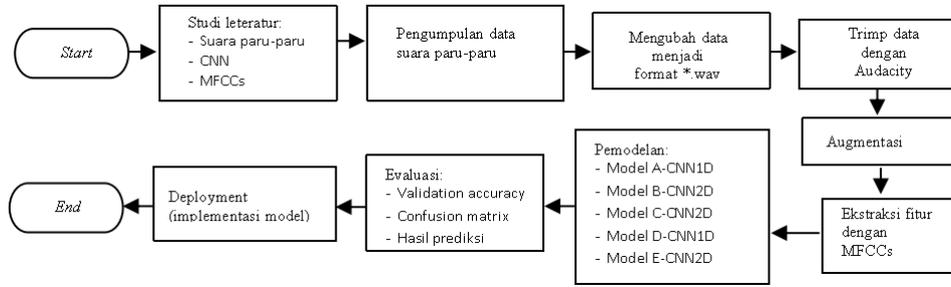
Dalam era kemajuan teknologi informasi dan komunikasi, pemrosesan sinyal suara menjadi semakin penting, terutama dalam aplikasi seperti pengenalan suara dan identifikasi pembicara. Salah satu metode yang telah terbukti efektif dalam pemrosesan suara adalah *Mel Frequency Cepstral Coefficient* (MFCC). MFCC sering digunakan sebagai teknik ekstraksi fitur dalam berbagai penelitian, termasuk sistem pengenalan suara, identifikasi, dan klasifikasi jenis suara [6]. Metode ini unggul karena kemampuannya menangkap informasi esensial dalam data suara dengan volume data yang minimal tanpa kehilangan informasi penting. Selain itu, pendekatan logaritmik yang digunakan dalam MFCC mencerminkan cara kerja sistem pendengaran manusia, menjadikannya sangat sesuai untuk analisis suara yang kompleks [11], [12]. Dalam konteks suara paru-paru, sinyal ini timbul akibat turbulensi udara selama proses pernapasan, terutama ketika udara mengalir dari saluran yang lebih lebar ke yang lebih sempit. Selama inspirasi, turbulensi ini lebih kuat sehingga menghasilkan suara yang lebih keras dibandingkan dengan ekspirasi. Suara paru-paru dapat dikategorikan ke dalam tiga jenis utama: normal, abnormal, dan tambahan, berdasarkan pitch, intensitas, lokasi, serta rasio inspirasi-ekspirasi [4].

Penelitian ini menggunakan MFCC sebagai teknik ekstraksi fitur utama karena kemampuannya menangkap karakteristik akustik penting dalam data suara. MFCC dikenal efektif untuk membedakan pola suara berdasarkan frekuensi yang meniru cara kerja sistem pendengaran manusia. Teknik ini sangat cocok untuk menganalisis suara pernapasan yang memiliki pola akustik kompleks. Di sisi lain, CNN merupakan algoritma yang andal untuk mengenali pola dan fitur visual, termasuk data audio yang telah diubah menjadi representasi berbasis spektrum. CNN dipilih karena kemampuannya dalam melakukan klasifikasi data berformat gambar dengan akurasi tinggi, menjadikannya solusi yang tepat untuk klasifikasi suara paru-paru. Untuk meningkatkan performa model, diterapkan berbagai teknik augmentasi data, seperti penambahan noise, time stretching, pitch shifting, dan random gain, yang bertujuan memperkaya variasi data pelatihan. Penelitian ini juga menerapkan metode *trial and error* untuk menemukan arsitektur CNN yang paling optimal.

Namun, terdapat gap yang belum teratasi dalam penelitian sebelumnya. Beberapa penelitian terdahulu, seperti yang dilakukan oleh [10], [11], [13], [14], belum mengintegrasikan teknik augmentasi data dan ekstraksi fitur MFCC ke dalam model CNN untuk klasifikasi suara paru-paru. Hal ini menyebabkan keterbatasan pada akurasi dan konsistensi model dalam mengenali berbagai jenis suara paru-paru. Perbedaan penelitian ini dengan sebelumnya terletak pada integrasi teknik augmentasi data dengan ekstraksi fitur MFCC untuk meningkatkan kinerja CNN secara signifikan. Penelitian ini bertujuan mengembangkan model klasifikasi suara paru-paru yang dapat mendeteksi gangguan pernapasan secara dini dan akurat. Diharapkan hasil penelitian ini dapat berkontribusi sebagai alat bantu diagnosis dibandingkan metode auskultasi manual.

## 2. METODE PENELITIAN

Metode penelitian ini terdiri dari beberapa langkah yang dilakukan secara bertahap sebagaimana pada Gambar 1. Langkah pertama adalah melakukan studi literatur terkait suara paru-paru, arsitektur CNN, dan teknik ekstraksi fitur MFCC. Selanjutnya, dilakukan proses *preprocessing* data, dimulai dari pengumpulan file audio yang semuanya berformat *.wav*, serta pemangkasan audio untuk setiap fase inspirasi dan ekspirasi. Proses ini dilakukan untuk sembilan kelas suara paru-paru yang akan digunakan sebagai dataset untuk pelatihan dan pengujian. Setelah itu, dataset dibagi menjadi data pelatihan dan data pengujian. Tahap pemodelan dilakukan dengan menguji lima arsitektur CNN yaitu Model A - CNN1D, Model B - CNN2D, Model C - CNN2D, Model D - CNN1D dan Model E - CNN2D. Kinerja setiap model dievaluasi menggunakan *confusion matrix* untuk menghitung akurasi dan menentukan model dengan performa terbaik.



Gambar 1. Tahapan penelitian

### 2.1. Pengumpulan dan Format Data

Dalam penelitian ini, data yang digunakan diperoleh dari beberapa sumber, yaitu *website database* ICBHI 2017, GitLab, dan saluran medis YouTube Medzcool dengan panjang sinyal paru-paru 5-30 detik yang diubah ke dalam format \*.wav. Total data sebanyak 1.350 sampel fase inspirasi dan ekspirasi pernapasan, yang terbagi ke dalam sembilan kelas suara paru-paru, yaitu *Krekels, Wheeze, Bronkial, Bronchovesicular, Vesicular, Stridor, mengi, Pleural Friction Rub, dan Trakeal*. Untuk mengatasi masalah pada data sebelum masuk ke proses pemodelan agar model yang dihasilkan berkualitas [15], [16]. Data penelitian dibagi menjadi 80:20, yaitu 80% untuk pelatihan dan 20% untuk pengujian serta validasi.

### 2.2. Trim Suara Paru-Paru

Rekaman suara paru-paru yang berdurasi antara 5 hingga 30 detik diproses menggunakan perangkat lunak Audacity. Setiap rekaman dipotong menjadi bagian inspirasi dan ekspirasi. Potongan tersebut disesuaikan dengan durasi 2 hingga 3 detik

### 2.3. Augmentasi Suara Paru-Paru

Augmentasi data adalah teknik yang digunakan untuk meningkatkan jumlah dan keragaman data pelatihan dalam *machine learning*, terutama dalam pembelajaran mendalam (*deep learning*) [17]. Teknik ini melibatkan berbagai transformasi pada data asli untuk menghasilkan data baru, yang dapat membantu model *machine learning* menjadi lebih *robust* dan generalisasi lebih baik. Pada proses augmentasi data suara paru-paru, digunakan beberapa teknik untuk memperkaya variasi data, seperti mengubah kecepatan dan *pitch*, menambahkan *noise*, serta memodifikasi durasi segmen suara dengan pemotongan atau pemanjangan, *time stretching, pitch shifting*, dan *random gain* [18].

### 2.4. Ekstraksi Data

Fitur ekstraksi yang digunakan pada penelitian ini adalah MFCC yang merupakan representasi dari nilai *real-cepstrum* dari sinyal suara melalui spektrogram dalam jangka waktu pendek [19], [20]. Proses ekstraksi MFCC diawali dengan langkah *framing*, yaitu memecah sinyal suara paru-paru menjadi bingkai-bingkai kecil berdurasi 25 ms dengan overlap 60%. Setelah itu, diterapkan *windowing* menggunakan jendela Hamming, sebagaimana pada Persamaan (1), untuk mengurangi distorsi spektrum di tepi *frame*.

$$h(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), n = 0, 1, 2, \dots, N-1 \tag{1}$$

Di mana  $h(n)$  adalah jendela Hamming,  $N$  adalah panjang frame dan  $n$  adalah indeks sampel.

Selanjutnya, Transformasi Fourier Diskrit (DFT) sebagaimana pada Persamaan (2) diterapkan untuk mengubah sinyal dari domain waktu ke domain frekuensi sehingga frekuensi dominan dalam sinyal dapat diidentifikasi.

$$S_i(k) = \sum_{n=1}^N S_i(n)h(n)e^{j2\pi kn/N}, 1 \leq k \leq K \tag{2}$$

Di mana  $S_i(k)$  adalah spektrum daya sinyal pada frekuensi ke- $k$ ,  $h(n)$  adalah jendela *Hamming* yang diterapkan pada sinyal,  $N$  adalah panjang DFT (jumlah titik dalam transformasi *Fourier*). Pada tahap *Mel-Frequency Wrapping*, spektrum frekuensi linier dari hasil DFT diubah ke skala Mel, yang lebih mencerminkan cara pendengaran manusia dengan frekuensi rendah yang lebih padat dan frekuensi tinggi yang lebih jarang. Konversi dari Hertz (Hz) ke Mel [19], [21] diberikan pada Persamaan (3).

$$Mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

Di mana  $f$  adalah frekuensi dalam satuan Mel. Tahap selanjutnya adalah aplikasi *Mel Filter bank* sesuai Persamaan (4) yang memisahkan energi spektral dalam setiap *frame* sinyal menjadi beberapa filter segitiga pada skala Mel untuk mencerminkan sensitivitas frekuensi pada pendengaran manusia.

$$E(i) = \sum_{k=1}^K |S_i(k)|^2 H_i(k) \tag{4}$$

Di mana  $E(i)$  adalah energi pada filter Mel ke- $i$ ,  $|S_i(k)|^2$  adalah magnitudo spektrum daya dari hasil DFT,  $H_i(k)$  adalah respon filter Mel ke- $i$  pada frekuensi ke- $k$ ,  $K$  adalah jumlah frekuensi dalam spektrum yang dianalisis. Untuk menyesuaikan persepsi manusia terhadap intensitas suara, diterapkan kompresi logaritmik pada energi dari filter Mel sebagaimana pada Persamaan (5).

$$E_{\log}(i) = \log(E(i)) \tag{5}$$

Dimana  $E_{\log}(i)$  adalah energi logaritmik pada filter Mel ke- $i$ . Kemudian terakhir, Discrete Cosine Transform (DCT) pada Persamaan (6) digunakan untuk mengubah spektrum energi logaritmik ke dalam domain waktu, yang menghasilkan koefisien MFCC [19]. Dimana  $c_k$  adalah koefisien cepstral ke- $k$ ,  $N$  adalah jumlah filter Mel.

$$c_k = \sum_{i=1}^N E_{\log}(i) \cos\left(\frac{\pi k(2i-1)}{2N}\right) \tag{6}$$

### 2.5. Pemodelan

Pemodelan menjelaskan langkah-langkah merancang arsitektur CNN. Klasifikasi menggunakan CNN dipilih karena memiliki tingkat akurasi yang sangat baik [22]. Pada perancangan ini menjelaskan bagaimana alur dari pelatihan model beserta pengujian model yang telah dibuat. Beberapa hal yang perlu dipersiapkan untuk klasifikasi CNN seperti membangun arsitektur, meng-*import Library* yang dibutuhkan, mengubah ukuran dataset. Proses pemodelan arsitektur dilakukan melalui beberapa percobaan dengan berbagai *hyperparameter* untuk menemukan model dengan akurasi terbaik. Lima jenis model yang diuji adalah Model A-CNN1D, Model B-CNN2D, Model C-CNN2D, Model D-CNN2D, dan Model E-CNN2D. Secara umum, arsitektur model CNN terdiri atas beberapa *layer*, yaitu *Convolutional layer*, *Pooling layer*, *Dropout layer*, *Flatten layer*, dan *Fully Connected layer*, seperti yang ditunjukkan pada Gambar 2 dan Gambar 3 untuk masing-masing jenis 1D dan 2D. Pada *Convolution Layer* digunakan input berukuran  $13 \times 9 \times 1$  dengan 16 filter, kernel  $3 \times 3$ , padding "same" (0), fungsi aktivasi ReLU, dan strides 1. Setelah konvolusi, proses dilanjutkan dengan *Pooling Layer*, *Dropout Layer*, *Flatten Layer*, dan *Fully Connected Layer* [22]. *Output layer* terdiri atas sembilan neuron dengan fungsi aktivasi Softmax. Fungsi Softmax mengonversi output menjadi probabilitas, di mana setiap nilai pada output menunjukkan probabilitas untuk masing-masing kelas.



Gambar 2. Representasi CNN 9 kelas suara paru model 1D



Gambar 3. Representasi CNN 9 kelas suara paru model 2D

### 2.6. Pelatihan, Prediksi dan Evaluasi Model

Pelatihan model CNN adalah tahap untuk melatih model dalam mendeteksi ekstraksi fitur dan klasifikasi suara paru-paru. Proses pelatihan dapat memakan waktu yang cukup lama. Jumlah *epoch* memengaruhi tingkat akurasi, tetapi *epoch* yang terlalu tinggi dapat memperpanjang waktu pelatihan. Evaluasi model dilakukan untuk menguji seberapa baik model mendeteksi dan mengklasifikasikan data suara paru-paru yang belum diketahui sebelumnya. Setelah melatih semua model, penilaian kinerja dilakukan dengan memeriksa rata-rata akurasi validasi dan titik konvergensi antara akurasi pelatihan dan validasi. Evaluasi dalam penelitian ini juga menggunakan *confusion matrix* untuk memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1 score* [13]. Berikut Persamaan *confusion matrix* (7), (8), (9), dan (10) yang digunakan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1. Augmentasi

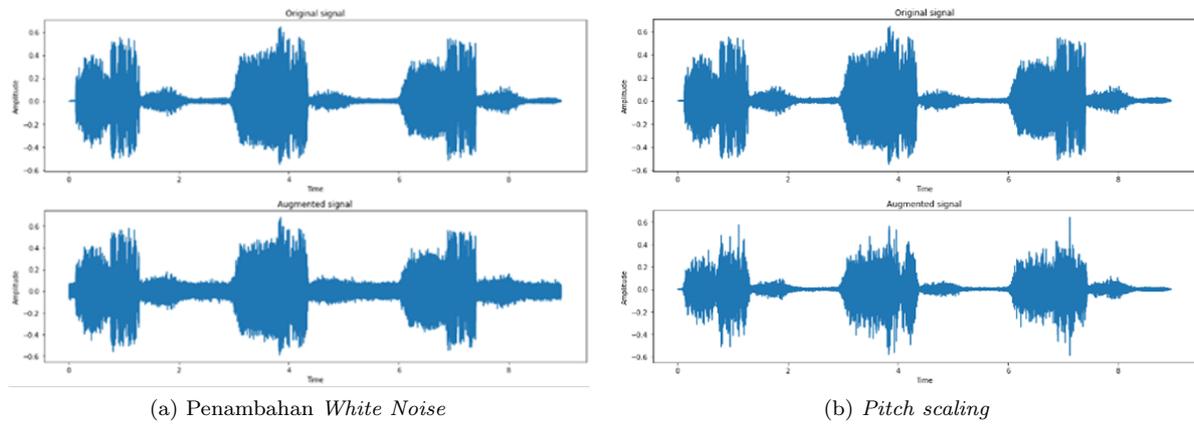
#### 3.1.1. Penambahan *White Noise*

Menambahkan *noise* ke data *audio* adalah salah satu teknik augmentasi yang sederhana namun efektif. *Noise* ditambahkan dengan menggabungkan data *audio* asli dengan derau acak yang dihasilkan oleh fungsi *np.random.randn()*. Parameter *noise\_percentage\_factor* mengontrol intensitas *noise* yang ditambahkan. Augmentasi ini membantu model untuk menjadi lebih *robust* terhadap gangguan *noise* di lingkungan nyata. Sampel penambahan *white noise* ditunjukkan pada Gambar 4a.

#### 3.1.2. *Pitch Scaling*

*Pitch calling* mengubah frekuensi dasar dari sinyal *audio* tanpa mengubah durasinya. Hal ini dilakukan dengan menggunakan fungsi *librosa.effects.pitch\_shift()*. Parameter *pitch\_factor* menentukan besarnya perubahan

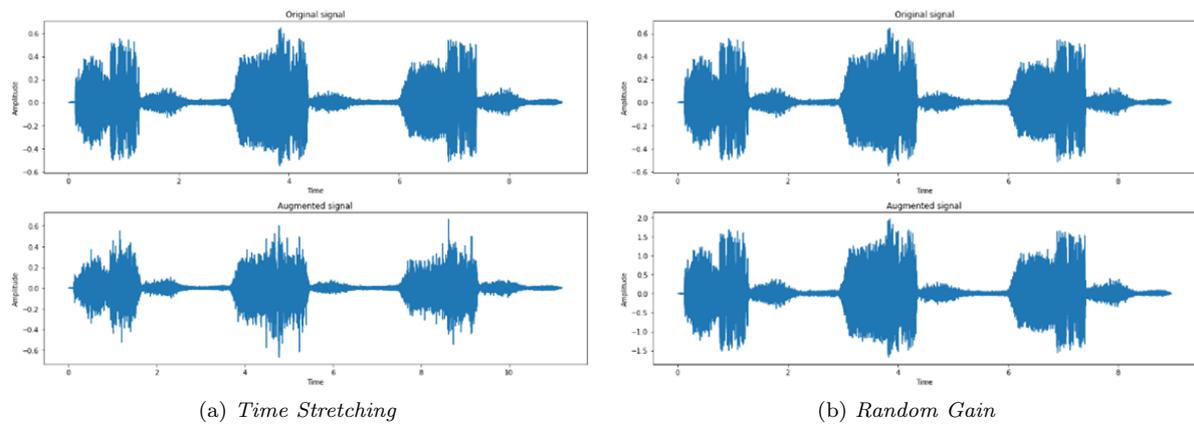
*pitch* yang diinginkan. Nilai positif akan menaikkan *pitch* sedangkan nilai negatif akan menurunkan *pitch*. Augmentasi ini membantu model mengenali variasi dalam suara pernapasan. Sampel penerapan *pitch scaling* ditunjukkan pada Gambar 4b.



Gambar 4. Augmentasi suara paru-paru dengan penambahan *White Noise* dan *Pitch Scaling*

### 3.1.3. Time Stretching

*Time stretching* mengubah durasi sinyal *audio* tanpa mengubah *pitch*. Dilakukan dengan menggunakan fungsi `librosa.effects.time_stretch()`. Parameter `stretch_factor` menentukan seberapa besar perubahan kecepatan yang diinginkan. Nilai lebih dari satu memperlambat *audio*, sementara nilai kurang dari satu mempercepat *audio*. Augmentasi ini membantu *model* untuk mengenali variasi dalam kecepatan pernapasan. Sampel penerapan *Time stretching* ditunjukkan pada Gambar 5a.



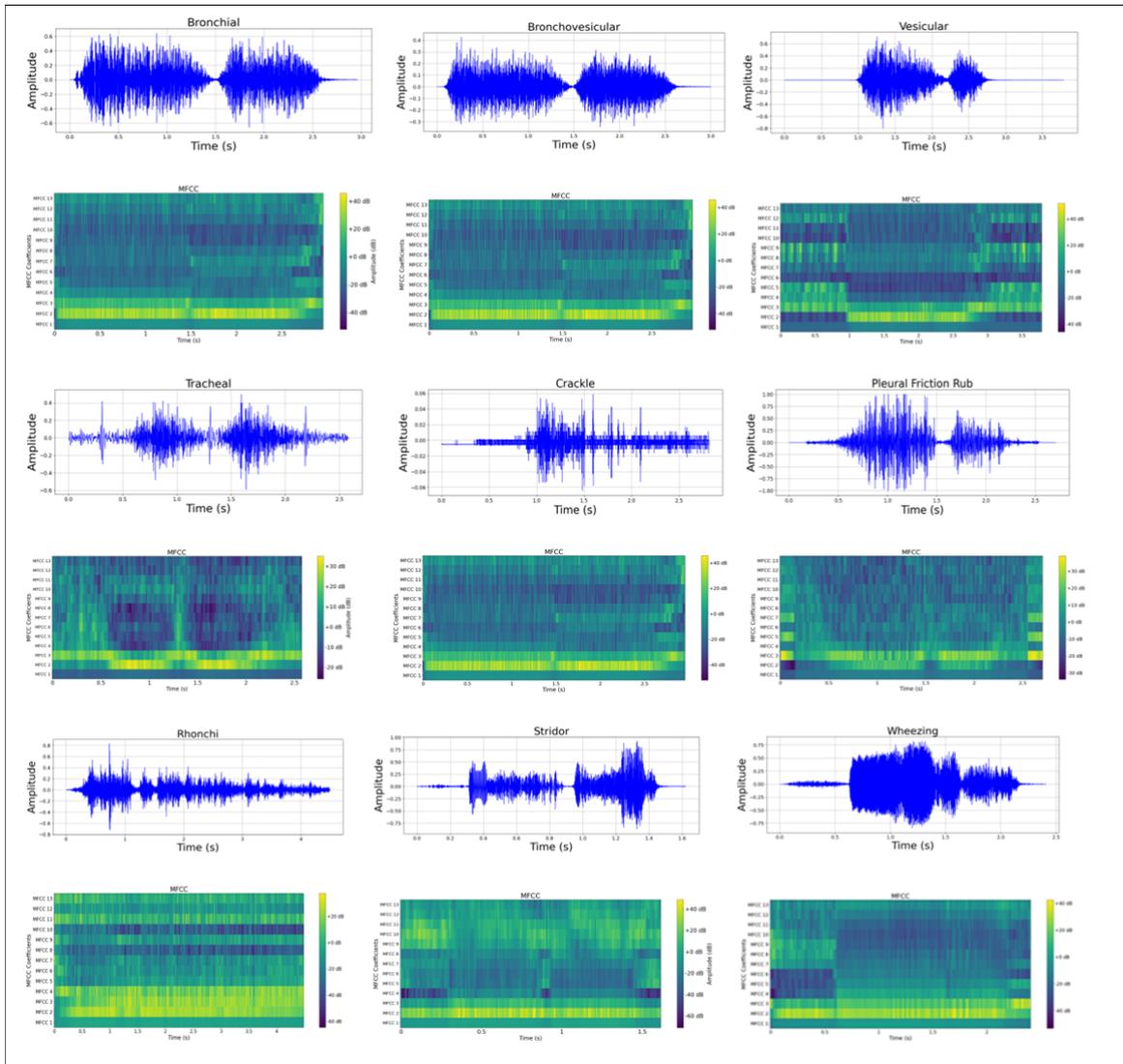
Gambar 5. Augmentasi data suara paru-paru dengan *Time Stretching* dan *Random Gain*

### 3.1.4. Random Gain

*Random gain* mengubah *amplitudo* sinyal *audio*, yaitu menambah atau mengurangi *volume*. Fungsi ini menggunakan `np.random.uniform(min_gain, max_gain)` untuk menghasilkan faktor pengali *gain* acak antara `min_gain` dan `max_gain`. Augmentasi ini membantu *model* untuk mengenali suara pernapasan pada berbagai tingkat *volume*. Sampel penerapan *random gain* ditunjukkan pada Gambar 5b.

### 3.2. Representasi Ekstraksi Fitur MFCC

MFCC pada Gambar 6 menunjukkan pola suara paru-paru terhadap waktu untuk sampel masing-masing sembilan kelas yaitu *Krekels*, *Wheeze*, *Bronkial*, *Bronchovesicular*, *Vesicular*, *Stridor*, *Ronki*, *Pleural Friction Rub*, dan *Trakeal*. Sumbu horizontal merepresentasikan waktu, sementara sumbu vertikal menggambarkan koefisien MFCC, yang mencerminkan frekuensi suara paru-paru. Setiap kelas memiliki pola koefisien yang unik, menandakan variasi suara paru-paru, di mana intensitas warna menunjukkan kekuatan energi suara paru-paru. Pola ini membantu mengidentifikasi perbedaan antara kondisi normal dan abnormal yang mencerminkan kondisi kesehatan paru-paru.



Gambar 6. Representasi MFCC untuk 9 kelas suara paru-paru

### 3.3. Analisis Pembentukan Model dengan CNN 1D dan 2D pada Data Suara Paru-Paru

Data suara paru-paru diklasifikasikan menggunakan algoritma CNN baik 1D maupun 2D dengan arsitektur dan *hyperparameter* yang berbeda di setiap model sebagaimana pada Tabel 1. Pemilihan CNN sebagai metode utama di sini sangat tepat mengingat CNN mampu menangkap pola spasial dan temporal dalam data, yang merupakan karakteristik penting dalam analisis sinyal suara medis.

Tabel 1. Pembentukan *Hyperparameter*

Model Percobaan	Hyperparameter				
	Model A-CNN1D	Model B-CNN2D	Model C-CNN2D	Model D-CNN1D	Model E-CNN2D
<b>Split data</b>		<b>80% Pelatihan, 10% Validasi, 10% Pengujian</b>			
Bathsize	32	32	32	32	32
Layers	18	8	11	19	12
Epoch	100	100	100	100	100
Optimizer	adam	adam	adam	adam	adam
Learning rate	0,01	0,01	0,01	0,01	0,01

### 3.3.1. Pembagian Data

Dataset dibagi menjadi 80% data untuk pelatihan, 10% untuk validasi, dan 10% untuk pengujian. Hal ini memastikan bahwa model mendapatkan cukup banyak data untuk belajar, sambil tetap menyisakan porsi data yang cukup untuk evaluasi performa model. Dengan dataset yang berjumlah 150 sampel, masing-masing subset data akan memiliki 120 sampel untuk pelatihan, 15 sampel untuk validasi, dan 15 sampel untuk pengujian.

### 3.3.2. Batch Size

Setiap model menggunakan *batch size* sebesar 32. *Batch size* yang digunakan konsisten untuk semua model (Model A sampai Model E), yang menunjukkan bahwa optimisasi dilakukan dalam *mini-batches*, di mana setiap iterasi pelatihan memperbarui bobot berdasarkan 32 sampel data. Pemilihan *batch size* ini seimbang antara waktu komputasi dan stabilitas pembelajaran, karena *batch size* yang lebih kecil sering kali memberikan gradien yang lebih fluktuatif, sementara yang terlalu besar bisa memerlukan komputasi yang lebih berat.

### 3.3.3. Arsitektur Model

Kelima model (A hingga E) memiliki perbedaan utama dalam jumlah layer konvolusi yang dipadukan dengan layer lain seperti *maxpooling*, *dropout*, *flatten*, dan *fully connected*. Model dengan lebih banyak layer, yakni Model A (18 layer) dan Model D (19 layer), dapat menangkap fitur lebih kompleks namun berisiko *overfitting* jika dataset tidak cukup besar. Sebaliknya, Model B (8 layer) dan Model C (11 layer) lebih sederhana dan cepat dilatih, meski mungkin kurang mampu menangkap fitur yang mendalam. Model E, dengan 12 layer, menawarkan keseimbangan antara kompleksitas dan efisiensi, mampu menangkap pola yang cukup detail tanpa risiko *overfitting* yang besar, model ini menjadi solusi optimal jika dataset tidak terlalu besar namun tetap membutuhkan pengenalan pola yang cukup mendalam.

### 3.3.4. Optimizer dan Learning rate

Kelima model menggunakan *optimizer Adam*, yang merupakan metode optimasi yang populer dalam *deep learning* karena kecepatan konvergensinya yang baik dan kemampuannya untuk menyesuaikan *learning rate* secara adaptif selama pelatihan. *Learning rate* ditetapkan sebesar 0,01, yang menandakan laju pembelajaran relatif cepat. *Learning rate* yang lebih tinggi cenderung mempercepat konvergensi namun bisa mengarah pada osilasi jika terlalu besar. Dalam konteks ini, pemilihan *learning rate* 0.01 dengan *optimizer Adam* adalah pilihan yang tepat untuk mencapai keseimbangan antara kecepatan konvergensi dan stabilitas pelatihan.

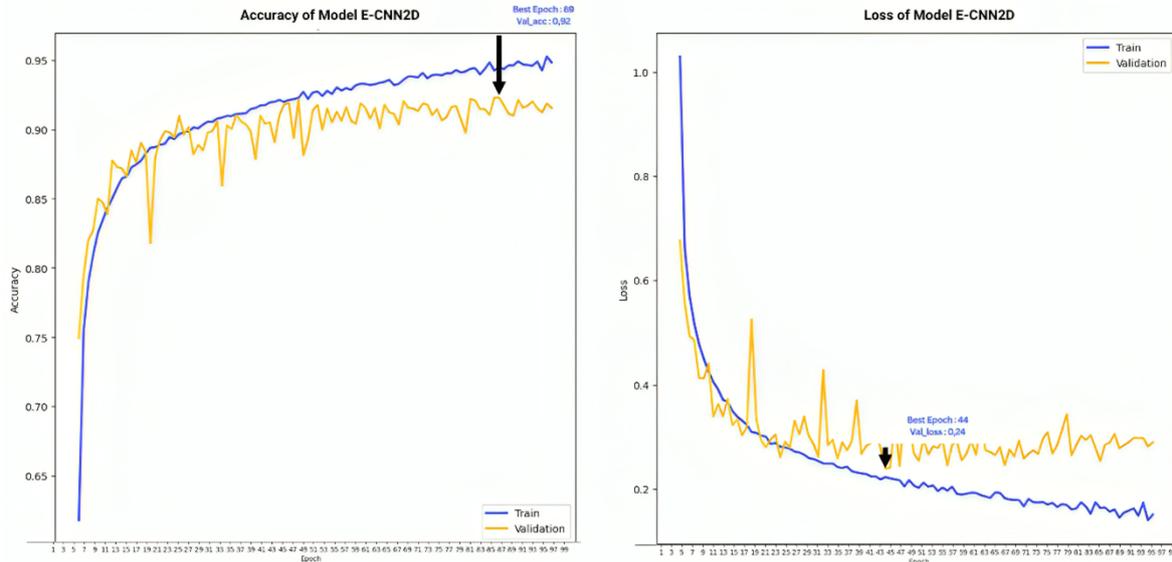
### 3.3.5. Epoch

Semua model dilatih selama 100 *epoch*. Jumlah ini menunjukkan banyaknya siklus penuh melalui seluruh dataset selama proses pelatihan. Dengan 100 *epoch*, model memiliki cukup kesempatan untuk mengeksplorasi dinamika pembelajaran tanpa risiko *overtraining*.

## 3.4. Pelatihan Arsitektur

Proses pelatihan data dilakukan untuk membangun *fitting model* yang akan digunakan untuk proses pengujian. Gambar 7 merupakan grafik dari salah satu model (E-CNN2D) yang menunjukkan pembelajaran

secara bertahap mempelajari data pelatihan. Hal ini terlihat dari kurva akurasi dan *loss* pada proses pelatihan dan validasi. Akurasi baik pada pelatihan maupun validasi menunjukkan peningkatan yang konsisten seiring dengan waktu, sedangkan nilai *loss* pada pelatihan dan validasi mengalami penurunan secara bertahap. Pola ini menunjukkan bahwa performa model dalam memprediksi semakin baik seiring bertambahnya iterasi. Nilai akurasi validasi terbaik tercapai pada iterasi ke-89 dengan akurasi sebesar 92%. Di sisi lain, nilai *loss* validasi terendah tercapai pada iterasi ke-44 dengan nilai *loss* sebesar 0,24%. Ini menunjukkan bahwa model tidak hanya semakin akurat dalam klasifikasi, tetapi juga mengalami penurunan kesalahan atau deviasi antara prediksi dan hasil sebenarnya.



Gambar 7. Grafik Akurasi dan *Loss* pada Data Pelatihan dan Validasi

### 3.5. Hasil Pengujian Model

Pengujian model dilakukan menggunakan dataset yang berjumlah 135 data. Setiap kelas dalam dataset tersebut terdiri dari 15 data suara paru-paru yang disimpan dalam format .csv. Tabel 2 menunjukkan sampel hasil pengujian untuk model E-CNN2D, yang berhasil mencapai akurasi sebesar 0,95.

Tabel 2. Beberapa Sampel Hasil Data Pengujian Model E-CNN2D

Fname	Label	Bronkial	Bronchoves	Krekels	Pleural	Ronki	Stridor	Trakeal	Vesicular	Mengi	Prediksi
6f57eb78.wav	Bronkial	0.948	3,83E+07	0.0	0.0	3,03E+09	3,08E+11	9,23E+10	0.048	0.003	Bronkial
9a67ca73.wav	Bronkial	0.995	1,79E+09	0.001	0.0	8,98E+08	0.0	7,76E+10	0.001	0.003	Bronkial
06f4ece6.wav	Bronchoves	0.002	0.887	0.0	0.009	2,02E+00	3,53E+09	0.017	0.0	0.085	Bronchovesicular
c884add3.wav	Bronchoves	0.008	0.901	0.001	0.082	5,00E+01	8,88E+08	0.008	0.001	0.0	Bronchovesicular
14b18ede.wav	Krekels	0.021	7,03E+10	0.732	0.005	2,89E+10	0.013	0.106	0.031	0.092	Krekels
4467b892.wav	Krekels	0.003	0.001	0.95	0.009	7,96E-07	0.0	0.004	0.018	0.015	Krekels
a99387ff.wav	Pleural	0.006	0.001	0.013	0.9	0.015	0.006	0.001	0.021	0.037	Pleural Friction
e1234aa4.wav	Pleural	8,65E+09	2,40E+09	4,44E+11	1.0	7,30E+04	1,16E+08	2,31E+11	2,64E+09	2,64E+10	Pleural Friction
7688e60d.wav	Ronki	3,49E+09	8,92E+04	2,14E+10	0.024	0.973	1,10E+10	3,14E-10	1,10E+04	0.003	Ronki
6be227b5.wav	Ronki	0.002	558	0.003	0.046	0.932	0.007	0.002	0.001	0.007	Ronki
68efb824.wav	Stridor	0.018	0.008	0.021	0.001	0.013	0.894	0.015	0.001	0.028	Stridor
c63b8198.wav	Stridor	0.013	0.012	0.005	0.057	0.076	0.754	0.01	0.001	0.072	Stridor
05d484a9.wav	Trakeal	0.001	0.034	0.002	0.099	1,04E+02	4,89E+08	0.861	0.002	0.001	Trakeal
0e45f12d.wav	Trakeal	0.008	0.001	0.017	0.004	0.003	0.243	0.338	0.026	0.36	Mengi
c3e7bf80.wav	Vesicular	0.008	103	0.001	0.026	2,65E+04	0.0	0.001	0.621	0.343	Vesicular
0754c120.wav	Vesicular	0.161	0.001	0.009	0.005	3,76E+04	0.001	0.003	0.783	0.038	Vesicular
00d70ff1.wav	Mengi	0.007	0.003	0.014	0.14	1,12E+09	0.002	0.004	0.004	0.827	Mengi
e6blea4.wav	Mengi	0.007	0.001	0.071	0.003	2,14E+01	0.0	0.0	0.002	0.916	Mengi

Accuracy Score: 0.9481481481481482

### 3.6. Beberapa Sampel Hasil Data Pengujian Model E-CNN2D

Untuk memastikan model perlu dilakukan pelatihan ulang atau tidak, dapat dilakukan dengan pengujian model menggunakan *confusion matrix* dengan output pengujian adalah nilai *accuracy*, *precision*, *recall* dan *f1*

score. Gambar 8 merupakan sampel hasil pengujian Model E-CNN2D dengan dataset sebanyak 135. Jenis suara paru-paru Bronkial terdapat *True Positive* (TP) sebesar 15 data, *True negative* (TN) sebanyak 120 data, *False Positive* (FP) sebesar 0 data, dan *False negative* (FN) sebanyak 0 data. Dengan menggunakan Persamaan (7), (8), (9) dan (10) diperoleh nilai *Accuracy* = 1, *precision* = 1, *recall* = 1, dan *F1-score* = 1. Jenis suara paru-paru yang lain dapat dilihat pada Tabel 3 dengan accuracy sebesar 0,95. hasil pengujian untuk semua arsitektur model ditunjukkan pada Tabel 4.



Gambar 8. Hasil *Confusion Matrix* Model E-CNN2D

Tabel 3. Hasil Pengujian Model E-CNN2D

Label	Precision	recall	f1-score
Bronkial	1,00	1,00	1,00
Bronchovesicular	1,00	0,80	0,89
Krekels	0,78	0,93	0,85
Pleural Friction	1,00	0,93	0,97
Ronki	0,94	1,00	0,97
Stridor	0,93	0,93	0,93
Trakeal	1,00	0,93	0,97
Vesicular	1,00	1,00	1,00
Mengi	0,94	1,00	0,97
Rata-rata	0,95	0,95	0,95
Accuracy	0,95		

Tabel 4. Hasil Pengujian Kelima Arsitektur Model

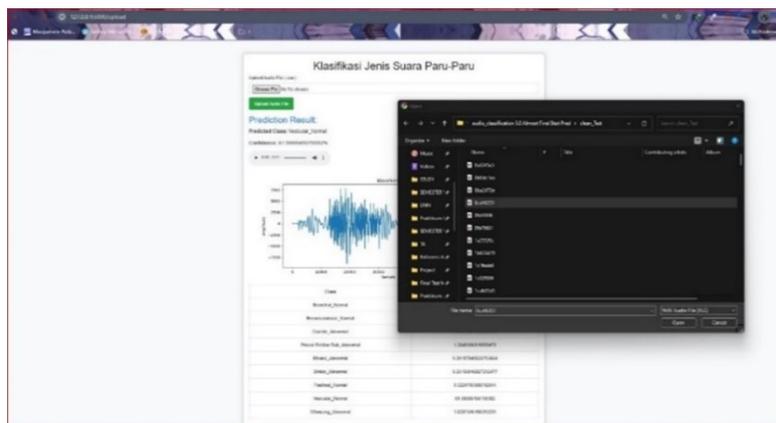
No	Model	Accuracy	Precision	Recall	F1-Score
1	Model A-CNN1D	0,83	0,92	0,73	0,81
2	Model B-CNN2D	0,93	0,95	0,93	0,94
3	Model C-CNN2D	0,92	0,92	0,92	0,92
4	Model D-CNN1D	0,84	0,88	0,84	0,85
5	Model E-CNN2D	0,95	0,95	0,95	0,95

Berdasarkan Tabel 4 yang menunjukkan hasil pengujian kelima jenis model, analisis metrik kinerja seperti *Accuracy*, *Precision*, *Recall* dan *F1-score*. Pada metrik *Accuracy*, Model E-CNN2D menampilkan akurasi tertinggi sebesar 0,95, diikuti oleh Model B-CNN2D dengan nilai 0,93. Sementara itu, Model A-CNN1D dan D-CNN1D memiliki akurasi lebih rendah, masing-masing sebesar 0,83 dan 0,84. Untuk metrik *Precision*, nilai tertinggi dicapai oleh Model E-CNN2D dan B-CNN2D dengan masing-masing mencapai 0,95. Meskipun Model A-CNN1D menunjukkan *precision* yang cukup tinggi sebesar 0,92, namun *recall* lebih rendah. Disisi lain, Model D-CNN1D memiliki *precision* paling rendah di antara semua, yaitu 0,88. Pada metrik *Recall*, Model

E-CNN2D kembali mencatat nilai tertinggi dengan 0,95, sama seperti precision-nya, menunjukkan konsistensi yang baik. Namun, Model A menunjukkan recall terendah dengan nilai 0,73, meskipun precision-nya tergolong tinggi. Selanjutnya, pada metrik *F1-Score*, Model E-CNN2D memiliki nilai *F1-Score* tertinggi, yaitu 0.95, yang menunjukkan keseimbangan baik antara *precision* dan *recall*. Sebaliknya, Model A-CNN1D memiliki *F1-Score* terendah, yaitu 0.81, mengindikasikan kurangnya keseimbangan antara *precision* dan *recall*.

### 3.7. Deployment/Implementasi

Hasil *deployment* berupa tampilan antarmuka dari model yang dihasilkan melalui pengolahan data menggunakan Python. Untuk membuat antarmuka pengujian model, proses *deployment* dilakukan menggunakan Framework Flask. Antarmuka yang dibuat memiliki fitur Prediksi Klasifikasi Jenis Suara Paru-Paru. Halaman ini digunakan untuk mengunggah file audio berformat \*.wav, yang kemudian akan diprediksi menggunakan model yang telah dilatih. Tampilan saat mengunggah dan melakukan prediksi suara paru-paru dapat dilihat pada Gambar 9.



Gambar 9. Halaman klasifikasi jenis suara paru-paru saat input dan prediksi suara paru-paru

Temuan penelitian ini menunjukkan bahwa model klasifikasi suara paru-paru yang menggunakan kombinasi teknik augmentasi data dengan ekstraksi fitur MFCC mampu secara signifikan meningkatkan kinerja model CNN dalam mengklasifikasikan suara paru-paru. Dengan teknik augmentasi seperti penambahan noise, perpanjangan waktu (*time stretching*), perubahan nada (*pitch shifting*), dan *random gain*, model E-CNN2D mencapai akurasi tertinggi sebesar 95% dalam mendeteksi sembilan kategori suara paru-paru, termasuk suara normal (bronkial, bronkovesikular, vesikular, trakeal) dan suara abnormal (mengi, stridor, ronki, krekels, pleural friction rub), mengungguli model-model lain yang mencapai akurasi antara 83% hingga 93%. Efektivitas penggabungan teknik augmentasi dan MFCC dalam menangkap informasi akustik penting memberikan kontribusi besar dalam meningkatkan kualitas pelatihan model, sehingga memungkinkan identifikasi pola suara paru-paru yang lebih akurat dan konsisten.

Penelitian ini juga memperlihatkan hasil yang lebih baik dibandingkan penelitian sebelumnya, seperti yang dilakukan oleh [10], [11], [13], dan [14], yang tidak mengoptimalkan penggabungan teknik augmentasi data dan MFCC dalam model CNN. Integrasi kedua teknik ini pada CNN menghasilkan model yang lebih andal sebagai alat bantu diagnosis dibandingkan metode auskultasi manual. Potensi besar model ini dalam mendukung proses diagnosis klinis secara cepat dan akurat relevan untuk diterapkan pada kasus gangguan pernapasan. Untuk meningkatkan kinerja lebih lanjut, penelitian ini merekomendasikan pengembangan pada model CNN2D, termasuk perluasan teknik augmentasi dan penggabungan jenis suara paru-paru tambahan agar prediksi model semakin optimal.

#### 4. KESIMPULAN

Pengembangan model klasifikasi suara paru-paru menggunakan kombinasi teknik augmentasi data, ekstraksi fitur MFCC dan CNN berhasil meningkatkan akurasi dalam mendeteksi gangguan pernapasan. Model E-CNN2D yang dirancang menunjukkan performa unggul dengan akurasi mencapai 95%, lebih tinggi dibandingkan model-model lainnya yang diuji dan penelitian sebelumnya. Teknik augmentasi data terbukti efektif dalam memperkaya variasi data latih dan meningkatkan kualitas pelatihan model, sehingga model mampu mengenali pola suara paru-paru secara akurat dan konsisten.

Penelitian ini juga menunjukkan bahwa integrasi augmentasi data dan MFCC pada CNN memberikan manfaat signifikan dalam klasifikasi suara paru-paru, menjadikannya alat bantu diagnosis yang lebih andal dibandingkan metode auskultasi manual. Dengan akurasi tinggi dan proses yang lebih cepat, model ini memiliki potensi besar untuk diterapkan dalam mendukung diagnosis klinis yang lebih efisien pada kasus gangguan pernapasan. Untuk meningkatkan kinerja model di masa depan, disarankan untuk memperluas teknik augmentasi dan memperkaya jenis suara paru-paru yang dianalisis, sehingga akurasi model dapat lebih dioptimalkan dalam berbagai kondisi klinis.

#### UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih sebesar-besarnya kepada tim ICBHI 2017 yang telah menyediakan database suara paru-paru yang sangat berharga untuk penelitian ini. Kami juga menghargai kontribusi dari GitLab sebagai platform pengelolaan dan kolaborasi kode yang memfasilitasi pengembangan model Kami. Ucapan terima kasih juga kami sampaikan kepada saluran YouTube Medzcool, yang menyediakan konten edukasi medis berkualitas dan membantu kami memperdalam pemahaman tentang analisis suara paru-paru.

#### DAFTAR PUSTAKA

- [1] W. Sun, Y. Zhang, dan F. Chen, "Research on heart and lung sound separation method based on DAE-NMF-VMD," *EURASIP Journal on Advances in Signal Processing*, vol. 2024, no. 1, p. 59, May 6, 2024. DOI: [10.1186/s13634-024-01152-0](https://doi.org/10.1186/s13634-024-01152-0).
- [2] O. Distler *et al.*, "Predictors of progression in systemic sclerosis patients with interstitial lung disease," *The European Respiratory Journal*, vol. 55, no. 5, p. 1902026, May 14, 2020. DOI: [10.1183/13993003.02026-2019](https://doi.org/10.1183/13993003.02026-2019). PMID: [32079645](https://pubmed.ncbi.nlm.nih.gov/32079645/).
- [3] X. Li *et al.*, "Research on electronic stethoscope system and signal processing algorithm," *Journal of Physics: Conference Series*, vol. 2634, no. 1, p. 012037, Nov. 1, 2023. DOI: [10.1088/1742-6596/2634/1/012037](https://doi.org/10.1088/1742-6596/2634/1/012037).
- [4] M. Fraiwan *et al.*, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, Apr. 2021. DOI: [10.1016/j.dib.2021.106913](https://doi.org/10.1016/j.dib.2021.106913).
- [5] Y. Kim *et al.*, "The coming era of a new auscultation system for analyzing respiratory sounds," *BMC Pulmonary Medicine*, vol. 22, no. 1, p. 119, Mar. 31, 2022. DOI: [10.1186/s12890-022-01896-1](https://doi.org/10.1186/s12890-022-01896-1).
- [6] L. Arts *et al.*, "The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: A meta-analysis," *Scientific Reports*, vol. 10, no. 1, p. 7347, Apr. 30, 2020. DOI: [10.1038/s41598-020-64405-6](https://doi.org/10.1038/s41598-020-64405-6).
- [7] J. S. Park *et al.*, "A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model," *Scientific Reports*, vol. 13, no. 1, p. 1289, Jan. 23, 2023. DOI: [10.1038/s41598-023-27399-5](https://doi.org/10.1038/s41598-023-27399-5).
- [8] H. Hafke-Dys *et al.*, "The Accuracy of Lung Auscultation in the Practice of Physicians and Medical Students," in *Advances in Medical Imaging, Detection, and Diagnosis*, 1st ed. New York: Jenny Stanford Publishing, Sep. 6, 2023, pp. 853–871. DOI: [10.1201/9781003298038-28](https://doi.org/10.1201/9781003298038-28).

- [9] F. Bertrand Z. *et al.*, “La auscultación pulmonar en el siglo 21,” *Revista Chilena de Pediatría*, vol. 91, no. 4, Aug. 14, 2020. DOI: [10.32641/rchped.v91i4.1465](https://doi.org/10.32641/rchped.v91i4.1465).
- [10] G. Petmezas *et al.*, “Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function,” *Sensors*, vol. 22, no. 3, p. 1232, Feb. 6, 2022. DOI: [10.3390/s22031232](https://doi.org/10.3390/s22031232).
- [11] D. Rafiqo, Y. Suyanto, dan C. Atmaji, “Klasifikasi Suara Paru-Paru Berdasarkan Ciri MFCC,” *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, vol. 12, no. 1, p. 1, Apr. 30, 2022. DOI: [10.22146/ijeis.70813](https://doi.org/10.22146/ijeis.70813).
- [12] H. Gulzar *et al.*, “Detection of Crackles and Wheezes in Lung Sound Using Transfer Learning,” *Health Informatics - An International Journal*, vol. 12, no. 2, pp. 01–14, May 27, 2023. DOI: [10.5121/hij.2023.12201](https://doi.org/10.5121/hij.2023.12201).
- [13] L. Nanni, G. Maguolo, dan M. Paci, “Data augmentation approaches for improving animal audio classification,” *Ecological Informatics*, vol. 57, p. 101 084, May 2020. DOI: [10.1016/j.ecoinf.2020.101084](https://doi.org/10.1016/j.ecoinf.2020.101084).
- [14] Y. Kim *et al.*, “Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning,” *Scientific Reports*, vol. 11, no. 1, p. 17 186, Aug. 25, 2021. DOI: [10.1038/s41598-021-96724-7](https://doi.org/10.1038/s41598-021-96724-7).
- [15] K. L.-M. Ang dan J. K. P. Seng, “Big Data and Machine Learning With Hyperspectral Information in Agriculture,” *IEEE Access*, vol. 9, pp. 36 699–36 718, 2021. DOI: [10.1109/ACCESS.2021.3051196](https://doi.org/10.1109/ACCESS.2021.3051196).
- [16] A. H. Al-Badri *et al.*, “Hybrid CNN Model for Classification of Rumex Obtusifolius in Grassland,” *IEEE Access*, vol. 10, pp. 90 940–90 957, 2022. DOI: [10.1109/ACCESS.2022.3200603](https://doi.org/10.1109/ACCESS.2022.3200603).
- [17] A. Agustina *et al.*, “Klasifikasi Penyakit Tanaman PADI Menggunakan Metode CNN Arsitektur Densenet-121 dan Augmentasi DataTA,” *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 8, no. 1, pp. 124–134, 1 Jun. 28, 2024. DOI: [10.35145/joisie.v8i1.4256](https://doi.org/10.35145/joisie.v8i1.4256).
- [18] M. M. E. Yahya, J. Jondri, dan A. Rizal, “Lung Sound Classification Using Wavelet Transformation and Recurrent Neural Networks,” in *E-Proceeding of Engineering*, vol. 8, Bandung: Telkom University, Apr. 2021, pp. 3209–3217.
- [19] Z. K. Abdul dan A. K. Al-Talabani, “Mel Frequency Cepstral Coefficient and its Applications: A Review,” *IEEE Access*, vol. 10, pp. 122 136–122 158, 2022. DOI: [10.1109/ACCESS.2022.3223444](https://doi.org/10.1109/ACCESS.2022.3223444).
- [20] J. Amose dan M. Vairavan, “Optimal chest position of auscultation for chronic obstructive pulmonary disease diagnosis using machine learning,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 3, p. 1417, Dec. 1, 2023. DOI: [10.11591/ijeecs.v32.i3.pp1417-1425](https://doi.org/10.11591/ijeecs.v32.i3.pp1417-1425).
- [21] A. rio Adriansyah, K. D. Prasetyo, dan H. A. A. A. Faruqi, “Pengenalan Pola Fonem Vokal menggunakan Short Time Fourier Transform (STFT) dan Fitur Mel Frequency Cepstral Coefficient (MFCC),” *Jurnal Teknologi Terpadu*, vol. 7, no. 1, pp. 1–6, 1 Jul. 16, 2021. DOI: [10.54914/jtt.v7i1.298](https://doi.org/10.54914/jtt.v7i1.298).
- [22] R. Hazra dan S. Majhi, “Detecting Respiratory Diseases from Recorded Lung Sounds by 2D CNN,” in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India: IEEE, Oct. 14, 2020, pp. 1–6. DOI: [10.1109/ICCCS49678.2020.9277101](https://doi.org/10.1109/ICCCS49678.2020.9277101).