Perancangan Pipeline Preprocessing dan Segmentasi Audio Batuk Berbasis Energi Adaptif untuk Klasifikasi Batuk Kering dan Berdahak

Desta Rizky Andhika[1], Faisal Muttaqin[2*], Fetty Tri Anggraeny [3]

[1]Informatika/Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
[1]22081010072@student.upnjatim.ac.id

[2]Informatika/Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
[2*]faisalmuttaqin.if@upnjatim.ac.id

[3]Informatika/Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
[3]fettyanggraeny.if@upnjatim.ac.id

Abstrak*— Rekaman audio batuk dari sumber publik umumnya memiliki variasi format, durasi, jumlah kanal, laju sampling, bagian hening, dan gangguan suara latar yang dapat memengaruhi konsistensi masukan pada proses klasifikasi. Penelitian ini bertujuan merancang pipeline preprocessing dan segmentasi audio batuk berbasis energi adaptif untuk menghasilkan audio yang lebih seragam dan relevan bagi klasifikasi batuk kering (dry) dan batuk berdahak (wet). Kebaruan penelitian ini terletak pada integrasi proses seleksi dan audit kualitas data dengan segmentasi event batuk menggunakan ambang energi yang dihitung secara adaptif pada setiap rekaman, sehingga proses segmentasi dapat menyesuaikan variasi intensitas sinyal antarrekaman tanpa menggunakan satu nilai ambang tetap untuk seluruh data. Tahapan penelitian meliputi seleksi label, audit kualitas audio, konversi format WAV, pengubahan kanal menjadi mono, resampling 16 kHz, trimming bagian hening, segmentasi berbasis energi adaptif, serta penyeragaman durasi menjadi 3 detik. Ambang segmentasi ditentukan secara adaptif berdasarkan estimasi noise floor, nilai RMS maksimum, dan margin sensitivitas sebesar 8.0 dB pada setiap rekaman.. Dari 27.550 entri metadata, diperoleh dataset akhir balanced-confident sebanyak 800 audio yang terdiri atas 400 data dry dan 400 data wet. Seluruh audio keluaran memiliki format WAV, kanal mono, laju sampling 16 kHz, dan durasi 3 detik. Hasil segmentasi menunjukkan rata-rata cakupan energi sebesar 84,58% pada kelas dry dan 82,54% pada kelas wet. Validasi fungsional menggunakan MFCC20 dan Capsule Network menghasilkan accuracy sebesar 71,67% dan F1-score sebesar 70,18%. Hasil tersebut menunjukkan bahwa pipeline mampu menghasilkan masukan audio yang konsisten dan dapat diteruskan ke tahap pemodelan klasifikasi.

Kata kunci— Audio batuk, Preprocessing audio, Segmentasi Energi Adaptif, MFCC, Capsule Network

Abstract*— Public cough audio recordings commonly vary in file format, duration, number of channels, sampling rate, silent segments, and background noise, which may affect the consistency of inputs used for classification. This study aims to design an adaptive energy-based cough audio preprocessing and segmentation pipeline to produce more standardized and relevant inputs for dry and wet cough classification. The novelty of this study lies in integrating data selection and audio quality auditing with cough-event segmentation using an energy threshold calculated adaptively for each recording, allowing the segmentation process to accommodate variations in signal intensity without applying a single fixed threshold to all recordings. The proposed stages include label selection, audio quality auditing, WAV conversion, mono-channel conversion, resampling at 16 kHz, silence trimming, adaptive energy-based segmentation, and duration standardization to 3 seconds. The segmentation threshold is adaptively determined for each recording based on the estimated noise floor, maximum RMS value, an 8.0 dB sensitivity margin. From 27,550 metadata entries, a final balanced-confident dataset of 800 recordings was obtained, consisting of 400 dry and 400 wet cough recordings. All output recordings were standardized into WAV format, mono channel, 16 kHz sampling rate, and 3-second duration. The segmentation results achieved average energy coverage of 84.58% for the dry class and 82.54% for the wet class. Functional validation using MFCC20 and Capsule Network achieved an accuracy of 71.67% and an F1-score of 70.18%. These results indicate that the proposed pipeline produces consistent audio inputs that can be further used for classification modeling.

Keywords— Cough Audio, Audio Preprocessing, Adaptive energy segmentation, MFCC, Capsule Network

I. PENDAHULUAN

Batuk merupakan salah satu gejala yang umum muncul pada gangguan pernapasan dan dapat direkam sebagai sinyal audio untuk dianalisis lebih lanjut. Karakteristik akustik pada suara batuk berpotensi dimanfaatkan dalam sistem klasifikasi untuk membedakan batuk kering (*dry cough*) dan batuk berdahak (*wet cough*). Namun, rekaman audio batuk dari sumber publik umumnya memiliki kondisi yang beragam, seperti perbedaan perangkat perekaman, durasi, intensitas suara, jumlah kanal, laju sampling, keberadaan bagian hening, serta noise

lingkungan. Variasi tersebut menyebabkan karakteristik sinyal antarrekaman tidak seragam dan dapat memengaruhi kualitas fitur yang dihasilkan pada proses klasifikasi[1], [2]

Pada pengolahan audio batuk, penggunaan sinyal mentah secara langsung berisiko membuat fitur yang diekstraksi tidak sepenuhnya merepresentasikan event batuk. Rekaman dapat mengandung bagian hening yang panjang, suara napas, suara percakapan, atau noise latar yang tidak relevan terhadap tujuan klasifikasi. Kondisi tersebut dapat menyebabkan proses ekstraksi fitur menangkap informasi non-batuk secara dominan, sehingga masukan yang diberikan kepada model klasifikasi menjadi kurang konsisten. Oleh karena itu, preprocessing audio diperlukan untuk menstandarkan karakteristik dasar sinyal dan mengurangi bagian rekaman yang tidak relevan sebelum tahap ekstraksi fitur.[3]

Sejumlah penelitian klasifikasi suara batuk telah memanfaatkan fitur akustik seperti *Mel-Frequency Cepstral Coefficients* (MFCC) dan model *deep learning* untuk mengenali pola pada audio medis. MFCC banyak digunakan karena mampu merepresentasikan karakteristik spektral suara berdasarkan skala Mel yang mendekati persepsi pendengaran manusia.[4] Sementara itu, Capsule Network (CapsNet) dapat digunakan untuk memproses representasi fitur audio melalui mekanisme *routing-by-agreement* yang mempertahankan hubungan antarfitur.[5] Meskipun demikian, keberhasilan tahap ekstraksi fitur dan klasifikasi tetap dipengaruhi oleh kualitas audio masukan. Oleh sebab itu, penyusunan pipeline preprocessing menjadi tahap penting sebelum data digunakan pada MFCC dan CapsNet.

Selain proses standarisasi format audio, segmentasi event batuk diperlukan untuk memperoleh bagian sinyal yang paling relevan. Segmentasi dilakukan karena satu rekaman audio tidak selalu hanya berisi satu event batuk, melainkan dapat memuat bagian hening, napas, suara lingkungan, atau aktivitas suara lain. Dalam penelitian deteksi batuk, proses segmentasi membantu membedakan bagian batuk dan non-batuk agar analisis selanjutnya lebih terfokus pada sinyal yang mengandung informasi akustik utama.[6] Pendekatan berbasis energi dapat digunakan untuk mengidentifikasi frame dengan aktivitas suara dominan. Namun, penggunaan ambang energi tetap berpotensi kurang optimal apabila seluruh rekaman menggunakan nilai ambang yang sama, karena intensitas suara batuk pada setiap audio dapat berbeda.

Berdasarkan permasalahan tersebut, penelitian ini merancang pipeline preprocessing dan segmentasi audio batuk berbasis energi adaptif. Pipeline mencakup proses seleksi label, audit kualitas audio, konversi format audio ke WAV, konversi stereo ke mono, resampling ke 16 kHz, trimming bagian hening, segmentasi event batuk berbasis energi adaptif, serta penyeragaman durasi audio menjadi 3 detik. Pada tahap segmentasi, nilai ambang energi ditentukan secara adaptif berdasarkan estimasi *noise floor* dan nilai RMS maksimum pada masing-masing rekaman dengan margin sensitivitas sebesar 8,0 dB. Dengan demikian, ambang dapat menyesuaikan tingkat energi dasar dan puncak sinyal setiap audio.[6]

Data penelitian menggunakan dataset publik *Cough Sound Dataset for Covid-19 Detection* yang memiliki metadata label batuk. Dari data yang tersedia, dilakukan seleksi terhadap label batuk kering dan batuk berdahak, pemeriksaan konflik label, verifikasi keberadaan file audio, serta audit kualitas rekaman. Proses tersebut menghasilkan dataset *balanced-confident* sebanyak 800 audio yang terdiri dari 400 batuk kering dan 400 batuk berdahak. Dataset yang seimbang digunakan agar setiap kelas memiliki proporsi yang sama pada proses pengolahan dan pengujian.[7]

Tujuan penelitian ini Adalah merancang pipeline preprocessing dan segmentasi audio batuk berbasis energi adaptif untuk menghasilkan data audio yang lebih seragam dan relevan sebagai masukan klasifikasi batuk kering dan berdahak. Kontribusi penelitian terletak pada penerapan proses seleksi data dan audit kualitas audio, segmentasi event batuk menggunakan ambang energi adaptif, serta penyeragaman keluaran audio menjadi format WAV mono dengan laju sampling 16 kHz dan durasi 3 detik.

Audio hasil pipeline selanjutnya diekstraksi menggunakan MFCC dan digunakan pada Capsule Network sebagai validasi pendukung bahwa keluaran preprocessing dapat diteruskan hingga tahap klasifikasi. Hasil klasifikasi tidak diposisikan sebagai diagnosis medis, melainkan sebagai bukti fungsional bahwa pipeline dapat menghasilkan data yang layak untuk pemodelan klasifikasi.

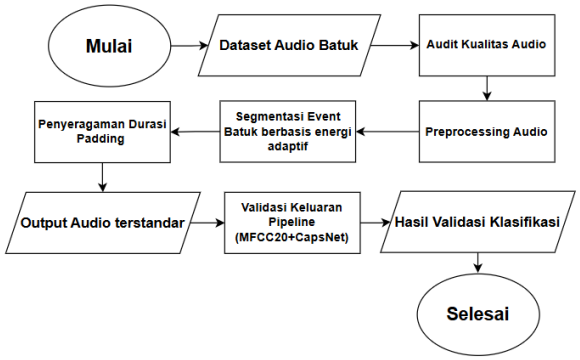
II. METODOLOGI PENELITIAN

A.Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan untuk menghasilkan audio batuk yang lebih konsisten sebelum digunakan pada proses klasifikasi. Tahapan dimulai dari seleksi data batuk kering dan batuk berdahak berdasarkan metadata yang tersedia. Data terpilih kemudian melalui audit kualitas untuk memastikan file audio dapat dibaca, memiliki durasi yang valid, dan mengandung aktivitas sinyal yang memadai.

Audio yang lolos audit diproses melalui konversi format apabila diperlukan, pengubahan kanal stereo menjadi mono, resampling menjadi 16 kHz, serta trimming bagian hening pada awal dan akhir rekaman. Selanjutnya, dilakukan segmentasi event batuk berbasis energi adaptif untuk memperoleh bagian audio dengan aktivitas energi dominan. Segmen hasil kemudian diseragamkan menjadi durasi 3 detik menggunakan padding atau cropping.

Keluaran pipeline berupa audio berformat WAV, mono, memiliki sampling rate 16 kHz, dan durasi seragam 3 detik. Audio tersebut selanjutnya digunakan pada ekstraksi fitur MFCC dan klasifikasi Capsule Network sebagai validasi pendukung bahwa hasil preprocessing dapat diteruskan ke tahap pemodelan



Gambar 1. Tahapan penelitian preprocessing dan segmentasi

B.Dataset dan Seleksi Data

Dataset penelitian berasal dari *Cough Sound Dataset for Covid-19 Detection* yang tersedia secara publik dan dilengkapi metadata. Dari dataset tersebut, penelitian hanya menggunakan rekaman yang memiliki label batuk kering (*dry*) dan batuk berdahak (*wet*), karena kedua kelas tersebut sesuai dengan tujuan klasifikasi pada penelitian ini. Data dengan label di luar kedua kategori tersebut tidak digunakan agar ruang lingkup penelitian tetap terarah.[7]

Seleksi data dilakukan sebelum rekaman memasuki tahap preprocessing. Proses ini mencakup pemeriksaan kesesuaian label pada metadata, identifikasi konflik label, serta verifikasi keberadaan file audio. Data yang memiliki indikasi label dry dan wet secara bersamaan tidak digunakan karena tidak dapat dipetakan secara konsisten ke dalam klasifikasi biner. Demikian pula, data yang tidak memiliki file audio yang sesuai tidak diteruskan ke tahap berikutnya.

Hasil seleksi awal menunjukkan bahwa distribusi data dry dan wet tidak seimbang, dengan jumlah rekaman batuk kering lebih besar daripada batuk berdahak. Oleh karena itu, setelah tahap audit kualitas audio, data valid digunakan untuk membentuk subset balanced-confident. Subset ini dibentuk dengan memilih jumlah data yang sama pada setiap kelas, yaitu 400 rekaman dry dan 400 rekaman wet, sehingga diperoleh dataset akhir sebanyak 800 audio. Dataset tersebut digunakan sebagai masukan pada pipeline preprocessing, segmentasi energi adaptif, penyeragaman durasi, dan validasi fungsional keluaran pipeline.

C.Audit Kualitas Audio

Audit kualitas audio dilakukan setelah tahap seleksi label untuk memastikan bahwa setiap rekaman yang digunakan memiliki kondisi teknis yang layak sebelum masuk ke pipeline preprocessing dan segmentasi. Pemeriksaan tidak hanya berfokus pada keberadaan file, tetapi juga memastikan audio dapat dibaca, hasil pemrosesan awal sesuai, serta memiliki aktivitas

sinyal yang memadai untuk dianalisis pada tahap segmentasi.

Pada tahap ini, setiap file diperiksa berdasarkan beberapa kriteria, yaitu keterbacaan file audio, kesesuaian file dengan metadata, keberadaan sinyal audio, hasil preprocessing, serta kelayakan kandidat event batuk. Rekaman yang gagal dibaca, tidak memiliki pasangan metadata yang valid, atau tidak memenuhi kriteria kelayakan sinyal tidak diteruskan ke tahap berikutnya. Langkah tersebut dilakukan untuk mencegah file bermasalah memengaruhi proses pembentukan dataset akhir.

Tabel II. Kriteria Audit Kualitas Audio

Kriteria	Pemeriksaan	Keputusan
Keterbacaan File	Audio dapat dibuka dan diproses	File gagal dibaca dikeluarkan
Kesesuaian metadata	File sesuai dengan entri metadata	Data tanpa metadata valid dikeluarkan
Aktivitas sinyal	Audio tidak kosong atau seluruhnya hening	Audio tanpa aktivitas memadai dikeluarkan
Hasil preprocessing	Parameter audio dapat diseragamkan	File bermasalah tidak diteruskan
Kelayakan segmentasi	Kandidat event dapat diidentifikasi	Audio tidak layak tidak digunakan

D.Preprocessing Audio

Preprocessing dilakukan untuk menyeragamkan karakteristik teknis rekaman sebelum proses segmentasi event batuk. Audio yang digunakan memiliki variasi format file, jumlah kanal, sampling rate, durasi, serta gangguan suara latar. Tahap preprocessing dilakukan agar keseluruhan data siap untuk pemodelan.[8] Perbedaan tersebut dapat menyebabkan hasil analisis sinyal antarrekaman tidak konsisten apabila audio langsung diproses tanpa tahap standardisasi. Tahap preprocessing diperlukan untuk mengurangi gangguan atau noise pada data masukan, sehingga proses klasifikasi dapat lebih terfokus pada karakteristik utama yang relevan, sebagaimana preprocessing pada penelitian sebelumnya digunakan untuk membantu model berfokus pada fitur penting setelah gangguan atau latar belakang yang tidak relevan dikurangi.[9] Preprocessing dilakukan melalui konversi format audio, konversi kanal menjadi mono, resampling, pengurangan noise, dan trimming bagian hening.[3], [10]

1) Konversi Format Audio

File audio dikonversi ke format WAV agar seluruh data memiliki format penyimpanan yang seragam. Penggunaan format WAV memudahkan proses pembacaan sinyal pada setiap file dan mengurangi perbedaan karakteristik pemrosesan akibat penggunaan format audio yang berbeda. Setelah tahap ini, seluruh audio diproses dalam format yang sama sebelum masuk ke tahap berikutnya.

2) Konversi Stereo ke Mono

Rekaman yang memiliki dua kanal audio diubah menjadi satu kanal mono. Tahap ini dilakukan karena penelitian berfokus pada karakteristik temporal dan spektral suara batuk, bukan pada perbedaan informasi antara kanal kiri dan kanan. Penggunaan audio mono juga membuat bentuk masukan lebih konsisten serta menyederhanakan proses segmentasi dan ekstraksi fitur[1], [10]. Sinyal mono diperoleh dengan merata-ratakan amplitudo kanal kiri dan kanan sebagai berikut:

$$x_{mono}[n] = \frac{x_L[n] + x_R[n]}{2} \qquad (1)$$

dengan $x_{mono}[n]$merupakan amplitudo sinyal mono pada sampel ke-n, sedangkan $x_L[n]$dan $x_R[n]$merupakan amplitudo pada kanal kiri dan kanan.

3) Resampling Audio

Seluruh audio diseragamkan pada sampling rate 16 kHz. Tahap ini diperlukan karena sampling rate awal setiap rekaman tidak selalu sama. Penyeragaman sampling rate memastikan bahwa setiap audio memiliki jumlah sampel per detik yang konsisten, sehingga proses pembagian frame, perhitungan energi, serta penyeragaman durasi dapat dilakukan menggunakan parameter yang sama.[1], [10]

4) Pengurangan Noise

Pengurangan noise dilakukan untuk menekan gangguan suara latar yang dapat berasal dari lingkungan perekaman, perangkat perekam, maupun aktivitas suara di sekitar subjek. Noise dapat menghasilkan komponen energi pada bagian audio yang tidak berkaitan dengan suara batuk dan berpotensi memengaruhi proses pemilihan kandidat segmen. Tahap ini digunakan sebagai proses pendukung untuk memperoleh sinyal yang lebih bersih sebelum dilakukan segmentasi berbasis energi.

5) Trimming Bagian Hening

Trimming dilakukan untuk mengurangi bagian hening pada awal dan akhir rekaman. Bagian hening umumnya memiliki amplitudo atau energi rendah dan tidak memberikan informasi utama dalam analisis suara batuk. Pengurangan bagian tersebut membantu proses segmentasi bekerja pada sinyal yang lebih terfokus terhadap aktivitas suara. Trimming pada tahap preprocessing tidak digunakan untuk menentukan batas akhir event batuk. Tahap ini hanya berfungsi sebagai penyaringan awal terhadap bagian audio yang tidak aktif, sedangkan penentuan kandidat event batuk dilakukan pada tahap segmentasi energi adaptif.

E.Segmentasi Event Batuk Berbasis Energi Adaptif

Setelah tahap preprocessing, rekaman masih dapat memuat suara napas, sisa noise latar, atau bagian suara lain yang tidak secara langsung merepresentasikan event batuk. Segmentasi dilakukan untuk memperoleh bagian audio dengan aktivitas akustik dominan sebelum proses penyeragaman durasi. Segmentasi event penting dilakukan karena batas kejadian batuk pada rekaman realistis dapat memengaruhi tahap analisis audio berikutnya[6]

1) Pembagian Sinyal ke dalam frame

Sinyal Audio hasil preprocessing dibagi menjadi beberapa frame pendek. Pembagian ini dilakukan agar perubahan aktivitas sinyal dapat dianalisis pada bagian waktu yang lebih kecil. Setiap frame kemudian dihitung nilai energinya untuk menunjukkan tingkat aktivitas audio pada frame tersebut.

Energi pada frame ke-i dihitung menggunakan persamaan:

$$E_i = \frac{1}{N}\sum_{n=1}^{N} x_i\,[n]^2 \qquad (2)$$

dengan E_i merupakan energi frame ke- i , $x_i[n]$adalah amplitudo sampel ke-n pada frame ke-i, dan N adalah jumlah sampel dalam satu frame.

Nilai energi yang lebih tinggi menunjukkan adanya aktivitas sinyal yang lebih dominan dibandingkan frame dengan energi rendah, seperti bagian hening atau suara latar berintensitas kecil.

2) Penentuan Ambang Energi Adaptif

Penggunaan satu nilai ambang tetap kurang sesuai untuk seluruh rekaman karena setiap audio memiliki tingkat kebisingan latar dan intensitas sinyal yang berbeda. Oleh karena itu, nilai ambang segmentasi ditentukan secara adaptif berdasarkan estimasi *noise floor* dan nilai RMS maksimum pada masing-masing rekaman. Pendekatan ini memungkinkan proses segmentasi menyesuaikan kondisi energi dasar setiap audio.[11] *Noise floor* diperoleh dari persentil ke-20 nilai RMS seluruh frame, sedangkan margin sensitivitas α ditetapkan sebesar 8,0 dB.

Untuk menyesuaikan tata letak dua kolom, perhitungan ambang dilakukan dalam dua tahap. Tahap pertama menentukan nilai ambang awal menggunakan Persamaan (2).

$$T_{\text{awal}} = \max(N_{\text{floor}} + \alpha,\, P_{\text{dB}} - 32) \qquad (3)$$

Tahap berikutnya membatasi nilai ambang awal terhadap puncak RMS menggunakan Persamaan (3).

$$T_{\text{dB}} = \min(T_{\text{awal}},\, P_{\text{dB}} - 1{,}5) \qquad (4)$$

dengan T_{awal} merupakan nilai ambang awal, T_{dB} merupakan nilai ambang segmentasi akhir, N_{floor} merupakan estimasi *noise floor*, P_{dB} merupakan nilai RMS maksimum seluruh frame, dan α merupakan margin sensitivitas sebesar 8,0 dB. Batas $P_{\text{dB}} - 32$digunakan agar nilai ambang tidak terlalu rendah terhadap puncak sinyal, sedangkan batas $P_{\text{dB}} - 1{,}5$ mencegah nilai ambang terlalu mendekati RMS maksimum. Karena *noise floor* dan RMS maksimum dihitung secara individual, nilai ambang yang dihasilkan dapat berbeda pada setiap rekaman sesuai karakteristik energinya.

3) Identifikasi Frame

Setelah nilai ambang segmentasi diperoleh, nilai RMS setiap frame dibandingkan dengan T_{dB}. Frame dengan nilai RMS sama dengan atau lebih besar dari ambang ditetapkan sebagai frame aktif, sedangkan frame dengan nilai RMS di bawah ambang ditetapkan sebagai frame tidak aktif. Proses identifikasi frame dirumuskan pada Persamaan (4).

$$a_i = \begin{cases} 1, & R_i \geq T_{\mathrm{dB}} \\ 0, & R_i < T_{\mathrm{dB}} \end{cases} \qquad (5)$$

dengan a_i merupakan status aktivitas frame ke-i, R_i merupakan nilai RMS frame ke-i, dan T_{dB} merupakan nilai ambang segmentasi. Nilai $a_i = 1$ menunjukkan bahwa frame memenuhi kriteria sebagai frame aktif, sedangkan nilai $a_i = 0$ menunjukkan bahwa frame berada di bawah ambang. Frame aktif yang muncul secara berurutan atau berdekatan selanjutnya digabungkan untuk membentuk kandidat segmen event batuk.

4) Penggabungan dan Pemilihan Segmen

Frame aktif yang berurutan atau berdekatan digabungkan menjadi kandidat segmen. Dari kandidat yang terbentuk, dipilih segmen dengan aktivitas energi paling dominan sebagai bagian audio utama. Tahap ini bertujuan mengurangi bagian rekaman yang tidak relevan, seperti hening panjang dan suara berenergi rendah, sehingga audio keluaran lebih terfokus pada aktivitas batuk.

Segmentasi ini tidak dimaksudkan sebagai penetapan diagnosis atau identifikasi klinis batuk secara langsung. Hasil segmentasi diposisikan sebagai kandidat bagian audio yang paling relevan untuk diteruskan ke tahap penyeragaman durasi dan ekstraksi fitur. Pendekatan ini selaras dengan fokus artikel, yaitu membangun pipeline audio yang lebih konsisten sebelum klasifikasi.

F.Penyeragaman Durasi Audio

Segmen audio hasil proses segmentasi memiliki durasi yang berbeda pada setiap rekaman. Perbedaan tersebut perlu diseragamkan agar seluruh audio memiliki bentuk masukan yang konsisten sebelum digunakan pada ekstraksi fitur dan validasi klasifikasi. Penyeragaman panjang sinyal diperlukan karena model klasifikasi berbasis audio memerlukan representasi fitur dengan ukuran yang seragam antarrekaman.[1]

Pada penelitian ini, durasi setiap audio ditetapkan menjadi 3 detik. Audio dengan durasi kurang dari 3 detik diberi *padding* berupa nilai nol pada bagian akhir sinyal. Sebaliknya, audio dengan durasi lebih dari 3 detik dipotong (*cropping*) dengan tetap mempertahankan segmen hasil segmentasi yang telah dipilih sebelumnya.

Jumlah sampel target dihitung menggunakan persamaan:

$$N_t = f_s \times d \qquad (6)$$

dengan N_t sebagai jumlah sampel target, f_s sebagai sampling rate, dan d sebagai durasi audio. Pada penelitian ini, sampling rate ditetapkan sebesar 16.000 Hz dan durasi target sebesar 3 detik, sehingga jumlah sampel akhir diperoleh sebagai sebagai berikut:

$$N_t = 16.000 \times 3 = 48.000 \text{ sampel} \qquad (7)$$

Penyeragaman durasi menghasilkan audio dengan panjang sinyal yang sama, yaitu 48.000 sampel. Kondisi tersebut membuat jumlah frame yang dihasilkan pada ekstraksi MFCC lebih konsisten dan memudahkan proses validasi menggunakan model klasifikasi tanpa dipengaruhi oleh variasi durasi antarrekaman

G.Validasi Keluaran Pipeline

Validasi keluaran pipeline dilakukan untuk memastikan bahwa audio hasil preprocessing, segmentasi energi adaptif, dan penyeragaman durasi dapat digunakan pada tahap ekstraksi fitur serta klasifikasi. Tahap ini tidak ditujukan untuk mengoptimalkan arsitektur model secara menyeluruh, melainkan untuk menguji kesiapan teknis audio keluaran pipeline sebagai masukan pemodelan.

Audio hasil pipeline diekstraksi menggunakan *Mel-Frequency Cepstral Coefficients* (MFCC) dengan konfigurasi 20 koefisien. MFCC digunakan untuk merepresentasikan karakteristik spektral suara batuk dalam bentuk fitur numerik yang dapat diproses oleh model klasifikasi. Selanjutnya, fitur tersebut digunakan sebagai masukan Capsule Network untuk membedakan batuk kering dan batuk berdahak.

Dataset final dibagi menggunakan metode *stratified split* dengan rasio 70:15:15, yang terdiri

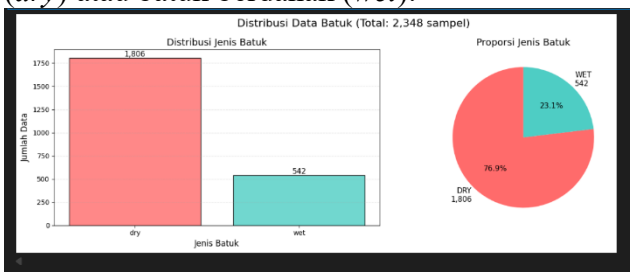
atas 560 data pelatihan, 120 data validasi, dan 120 data pengujian. Pembagian secara stratified digunakan agar proporsi kelas batuk kering dan batuk berdahak tetap seimbang pada setiap subset.

Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score* untuk memberikan gambaran umum bahwa audio keluaran pipeline dapat diproses hingga tahap klasifikasi. Hasil evaluasi tidak digunakan untuk membandingkan skenario model maupun mengoptimalkan arsitektur Capsule Network, melainkan sebagai validasi fungsional bahwa preprocessing, segmentasi energi adaptif, dan penyeragaman durasi menghasilkan masukan yang dapat digunakan pada proses pemodelan.

III. HASIL DAN PEMBAHASAN

A.Hasil Seleksi Data dan Audit Kualitas Audio

Seleksi data dan audit kualitas audio dilakukan untuk memperoleh rekaman yang memiliki label sesuai serta kondisi teknis yang layak sebelum diproses pada pipeline. Dataset awal terdiri atas 27.550 entri metadata. Dari jumlah tersebut, sebanyak 2.348 data memiliki label batuk kering (*dry*) atau batuk berdahak (*wet*).



Gambar 2. Distribusi awal data batuk berlabel dry dan wet sebelum penanganan konflik label.

Pada tahap verifikasi label, ditemukan 44 data yang memiliki indikasi label *dry* dan *wet* secara bersamaan. Data dengan label ganda tersebut tidak digunakan karena tidak dapat dipetakan secara konsisten ke dalam klasifikasi biner. Setelah konflik label dikeluarkan, diperoleh 2.188 rekaman yang digunakan sebagai masukan pada tahap audit kualitas audio.

Audit dilakukan untuk memastikan file audio dapat dibaca dan memiliki kondisi sinyal yang memadai untuk melalui tahapan preprocessing serta segmentasi. Hasil audit menunjukkan bahwa 2.177 rekaman dinyatakan valid dan dapat diproses lebih lanjut. Rekaman yang tidak lolos audit tidak digunakan agar pipeline tidak dipengaruhi oleh file yang berpotensi menghasilkan keluaran tidak representatif.

Data valid kemudian digunakan untuk membentuk subset *balanced-confident*. Dataset akhir terdiri atas 800 rekaman, yaitu 400 batuk kering dan 400 batuk berdahak. Pembentukan data yang seimbang dilakukan untuk mengurangi dominasi kelas mayoritas dan menjaga proses validasi klasifikasi tetap proporsional. Dataset akhir tersebut selanjutnya digunakan pada preprocessing audio, segmentasi event batuk berbasis energi adaptif, penyeragaman durasi, dan validasi fungsional keluaran pipeline

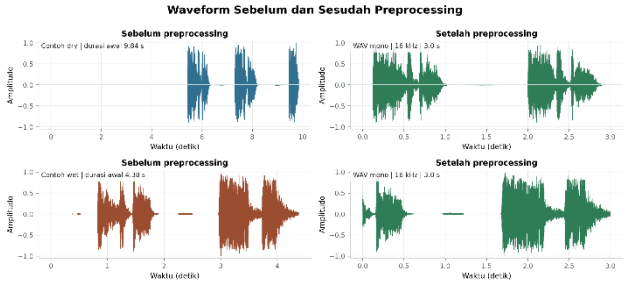
Tabel IV. Ringkasan Seleksi, Audit, dan Pembentukan Dataset

Tahap	Jumlah Data	Keterangan
Metadata Awal	27.550	Basis metadata penelitian
Data berlabel dry/wet	2.348	Termasuk data dengan konflik label
Konflik label dikeluarkan	44	Tidak diteruskan ke audit kualitas
Lolos seleksi label	2.188	Masukkan audit kualitas audio
Audio valid	2.177	Layak diproses lebih lanjut
Dataset akhir	800	400 dry dan 400 wet

B.Hasil Preprocessing dan Segmentasi Energi Adaptif.

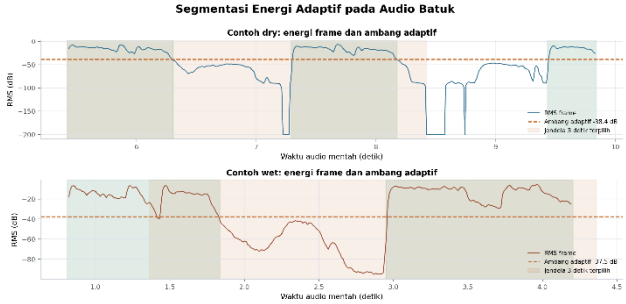
Hasil preprocessing menunjukkan bahwa setiap rekaman berhasil diseragamkan menjadi audio berformat WAV, satu kanal mono, sampling rate 16 kHz, dan durasi akhir 3 detik. Gambar 3 memperlihatkan perubahan bentuk sinyal pada contoh batuk *dry* dan *wet*. Sebelum preprocessing, rekaman masih memiliki durasi yang berbeda, yaitu 9,84 detik pada contoh *dry* dan 4,38 detik pada contoh *wet*. Setelah preprocessing, kedua audio memiliki durasi sama,

yaitu 3 detik, sehingga bentuk masukan menjadi lebih konsisten untuk tahap berikutnya.



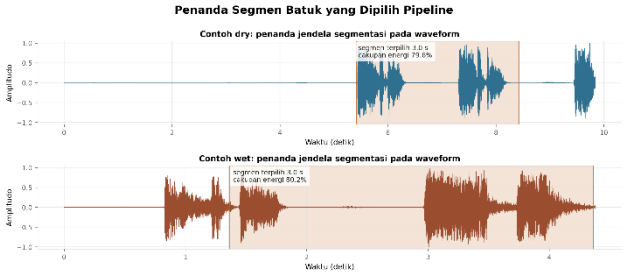
Gambar 3. Waveform audio sebelum dan sesudah preprocessing pada contoh batuk dry dan wet

Selain penyeragaman durasi, preprocessing juga mengurangi bagian hening pada rekaman. Pada waveform hasil preprocessing, aktivitas sinyal utama terlihat lebih terfokus dibandingkan waveform awal. Namun, tahap ini tidak dimaksudkan untuk menentukan batas event batuk secara akhir. Penentuan kandidat segmen tetap dilakukan melalui analisis energi frame pada tahap segmentasi adaptif.



Gambar 4. Energi frame dan Ambang Adaptif

Gambar 4 menunjukkan distribusi nilai RMS setiap frame dan ambang energi adaptif pada contoh rekaman batuk kering (*dry*) dan batuk berdahak (*wet*). Garis putus-putus menunjukkan nilai ambang yang dihitung secara terpisah pada setiap rekaman berdasarkan estimasi *noise floor*, nilai RMS maksimum, dan margin sensitivitas sebesar 8,0 dB. Pada contoh *dry*, diperoleh nilai ambang sebesar −38,4 dB, sedangkan pada contoh *wet* sebesar −37,5 dB. Perbedaan nilai tersebut terjadi karena setiap rekaman memiliki tingkat energi dasar dan puncak RMS yang berbeda. Dengan demikian, nilai ambang tidak ditetapkan secara tetap untuk seluruh data, tetapi menyesuaikan karakteristik energi masing-masing rekaman. Frame dengan nilai RMS sama dengan atau lebih besar dari ambang ditetapkan sebagai frame aktif dan selanjutnya digunakan untuk membentuk kandidat segmen event batuk.



Gambar 5. Jendela segmentasi berdurasi 3 detik yang dipilih pada waveform audio batuk dry dan wet

Gambar 5 menunjukkan jendela segmentasi berdurasi 3 detik yang dipilih dari rekaman awal. Pada contoh *dry*, segmen terpilih memiliki cakupan energi sebesar 79,6%, sedangkan pada contoh *wet* mencapai 80,2%. Nilai tersebut menunjukkan bahwa jendela yang dipilih mencakup sebagian besar aktivitas energi dominan pada rekaman. Dengan demikian, proses segmentasi dapat mengurangi bagian hening atau aktivitas berenergi rendah tanpa menghilangkan bagian utama sinyal yang digunakan sebagai masukan klasifikasi.

Secara keseluruhan, hasil ringkasan segmentasi menunjukkan bahwa proses dapat diterapkan pada seluruh 2.188 audio hasil seleksi tanpa menggunakan mekanisme *fallback*. Rata-rata jumlah kandidat event pada kelas *dry* adalah 3,11 event dan pada kelas *wet* sebesar 3,25 event. Dari kandidat tersebut, rata-rata event yang terpilih masing-masing sebanyak 2,12 dan 2,14 event. Kelas *dry* memiliki rata-rata cakupan energi sebesar 84,58%, sedangkan kelas *wet* sebesar 82,54%. Hasil tersebut menunjukkan bahwa segmentasi adaptif mampu memilih bagian audio yang mencakup aktivitas energi dominan pada kedua kelas. Ringkasan segmentasi pada Tabel V dihitung pada 2.188 rekaman hasil seleksi label. Sementara itu, verifikasi karakteristik output akhir dan validasi klasifikasi dilakukan pada subset balanced-confident sebanyak 800 audio.

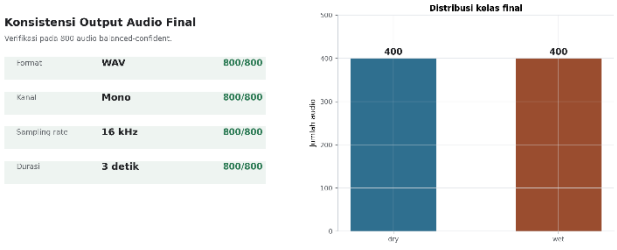
Tabel V. Ringkasan Hasil Segmentasi Energi Adaptif

Kelas	Jumlah Audio	Rata-rata Kandidat Event	Rata-rata event terpilih	Rata-rata cakupan Energi
Dry	1.678	3,11	2,12	84,58%
Wet	510	3,25	2,14	82,54%

C.Hasil Penyeragaman Output Audio

Penyeragaman output dilakukan pada dataset akhir *balanced-confident* yang terdiri atas 800 audio, dengan komposisi 400 rekaman batuk kering dan 400 rekaman batuk berdahak. Verifikasi terhadap keluaran pipeline menunjukkan bahwa seluruh audio memiliki karakteristik teknis yang konsisten, yaitu format WAV, satu kanal mono, sampling rate 16 kHz, dan durasi 3 detik.

Konsistensi tersebut menunjukkan bahwa tahapan konversi format, pengubahan kanal, resampling, segmentasi, serta padding atau *cropping* berhasil menghasilkan bentuk masukan yang seragam. Dengan sampling rate 16 kHz dan durasi 3 detik, setiap rekaman memiliki panjang sinyal sebesar 48.000 sampel. Keseragaman panjang sinyal ini penting agar ekstraksi MFCC menghasilkan representasi fitur dengan dimensi yang konsisten pada seluruh data.



Gambar 6. Karakteristik Output Audio dan Distribusi kelas dataset akhir

Gambar 6 memperlihatkan bahwa seluruh 800 audio memenuhi empat parameter output yang ditetapkan. Selain itu, distribusi kelas akhir tetap seimbang, yaitu 400 data *dry* dan 400 data *wet*. Dengan demikian, keluaran pipeline tidak hanya seragam secara teknis, tetapi juga mempertahankan komposisi kelas yang proporsional untuk tahap validasi klasifikasi.

D.Validasi Fungsional Keluaran Pipeline

Validasi fungsional dilakukan untuk memastikan bahwa audio hasil preprocessing, segmentasi energi adaptif, dan penyeragaman durasi dapat digunakan sebagai masukan pada tahap ekstraksi fitur dan klasifikasi. Pada tahap ini, audio keluaran pipeline diekstraksi menggunakan 20 koefisien *Mel-Frequency Cepstral Coefficients* (MFCC20), kemudian diproses menggunakan Capsule Network untuk membedakan batuk kering dan batuk berdahak.

Dataset final dibagi secara *stratified* menjadi 560 data pelatihan, 120 data validasi, dan 120 data pengujian. Representasi MFCC20 pada data pelatihan memiliki bentuk (560, 299, 20, 3) yang menunjukkan bahwa seluruh audio keluaran pipeline dapat dibentuk menjadi representasi fitur dengan dimensi yang konsisten. Konsistensi ini menunjukkan bahwa standardisasi format, kanal, sampling rate, dan durasi telah mendukung proses ekstraksi fitur tanpa kendala bentuk masukan.

Hasil pengujian pada data uji menunjukkan bahwa konfigurasi MFCC20 dan Capsule Network memperoleh *accuracy* sebesar 71,67%, *precision* 74,07%, *recall* 66,67%, dan *F1-score* 70,18%. Nilai *accuracy* menunjukkan bahwa model mampu mengklasifikasikan 71,67% data uji secara benar. Sementara itu, *F1-score* merepresentasikan keseimbangan antara *precision* dan *recall*, sehingga evaluasi tidak hanya mempertimbangkan jumlah prediksi yang benar, tetapi juga ketepatan prediksi dan kemampuan model dalam mengenali kelas positif.[12] Nilai *precision* yang lebih tinggi daripada *recall* menunjukkan bahwa prediksi positif yang dihasilkan relatif tepat, meskipun masih terdapat sebagian data positif yang belum berhasil dikenali. Hasil tersebut tidak digunakan untuk menyatakan bahwa performa model telah memenuhi standar diagnostik, melainkan sebagai validasi fungsional bahwa audio hasil *preprocessing* dan segmentasi masih mempertahankan karakteristik akustik yang dapat dipelajari oleh model klasifikasi.

Tabel VI. Hasil Validasi Fungsional Keluaran Pipeline

Konfigurasi	Accuracy	F1-Score

MFCC20 + Capsule Network		71,67%	70,18%

IV. KESIMPULAN

Penelitian ini berhasil merancang pipeline preprocessing dan segmentasi audio batuk berbasis energi adaptif untuk menghasilkan data audio yang lebih seragam sebelum klasifikasi batuk kering dan batuk berdahak. Melalui seleksi label, audit kualitas, preprocessing, segmentasi energi adaptif, serta penyeragaman durasi, diperoleh dataset *balanced-confident* sebanyak 800 audio yang seluruhnya memiliki format WAV, kanal mono, sampling rate 16 kHz, dan durasi 3 detik. Segmentasi mampu memilih bagian sinyal dengan aktivitas energi dominan, dengan rata-rata cakupan energi sebesar 84,58% pada kelas *dry* dan 82,54% pada kelas *wet*. Validasi menggunakan MFCC20 dan Capsule Network menghasilkan *accuracy* 71,67% serta *F1-score* 70,18%, yang menunjukkan bahwa keluaran pipeline dapat digunakan sebagai masukan klasifikasi. Hasil tersebut menegaskan bahwa pipeline yang dirancang mampu meningkatkan konsistensi teknis audio dan menyediakan representasi masukan yang layak untuk tahap pemodelan. Meskipun demikian, penelitian ini masih terbatas pada penggunaan label yang bersumber dari metadata dataset publik, data yang berasal dari satu sumber, serta pendekatan segmentasi berbasis RMS yang belum secara langsung membedakan event batuk dari suara nonbatuk berenergi tinggi. Oleh karena itu, penelitian selanjutnya dapat diarahkan pada validasi label oleh tenaga ahli, pengujian menggunakan dataset eksternal dengan kondisi perekaman yang lebih beragam, serta pengembangan metode segmentasi yang mengombinasikan karakteristik energi dan spektral untuk meningkatkan ketepatan identifikasi event batuk.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur atas dukungan akademik yang diberikan selama proses pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada dosen pembimbing yang telah memberikan arahan, masukan, dan evaluasi dalam penyusunan penelitian. Selain itu, penulis mengucapkan terima kasih kepada penyedia dataset publik COUGHVID yang telah menyediakan data audio batuk sehingga penelitian mengenai pipeline preprocessing dan segmentasi audio batuk berbasis energi adaptif ini dapat dilaksanakan.

REFERENSI

[1] M. Pahar *et al.*, "Automatic Tuberculosis and COVID-19 cough classification using deep learning," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Jul. 2022, pp. 1–9. doi: 10.1109/ICECET55527.2022.9873469.

[2] A. Muhammad, M. A. Arserim, and Ö. Türk, "Compare the classification performances of convolutional neural networks and capsule networks on the Coswara dataset," *Dicle University Journal of Engineering*, vol. 14, no. 2, pp. 265–271, 2023, doi: 10.24012/dumf.1270429.

[3] A. Ö. TürkçetiN, T. Koç, and Ş. ÇiLekar, "COUGH SOUND ANALYSIS WITH DEEP LEARNING: THE IMPACT OF DATA AUGMENTATION ON RESPIRATORY DISEASE CLASSIFICATION," *Mühendislik Bilimleri ve Tasarım Dergisi*, vol. 13, no. 3, pp. 896–910, 2025, doi: 10.21923/jesd.1672180.

[4] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980, doi: 10.1109/TASSP.1980.1163420.

[5] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," Nov. 07, 2017, *arXiv*: arXiv:1710.09829. doi: 10.48550/arXiv.1710.09829.

[6] M. You, W. Wang, Y. Li, J. Liu, X. Xu, and Z. Qiu, "Automatic cough detection from realistic audio recordings using C-BiLSTM with boundary regression," *Biomedical Signal Processing and Control*, vol. 72, p. 103304, Feb. 2022, doi: 10.1016/j.bspc.2021.103304.

[7] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Sci Data*, vol. 8, no. 1, p. 156, Jun. 2021, doi: 10.1038/s41597-021-00937-4.

[8] C. Ramadhan, V. Atina, and H. Permatasari, "Analisis Perbandingan Model CNN dan IndoBERT Dalam Sentimen Berita Politik Indonesia," pp. 110–118, 2025, doi: 10.47701/v1r9ka69.

[9] S. Afriza, N. A. C. Pratama, M. A. Aziz, and F. Muttaqin, "Implementation of CNN Method with Otsu Thresholding Preprocessing for Pneumonia Detection," *JITTER: Jurnal Ilmiah Teknologi dan Komputer*, vol. 7, no. 1, 2026.

[10] K. M. Rezaul, "Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, 2024.

[11] E. P. Mandyartha, F. T. Anggraeny, F. Muttaqin, and F. A. Akbar, "Global and Adaptive Thresholding Technique for White Blood Cell Image Segmentation," *J. Phys.: Conf. Ser.*, vol. 1569, no. 2, p. 022054, Jul. 2020, doi: 10.1088/1742-6596/1569/2/022054.

[12] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci Rep*, vol. 12, no. 1, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.