

Klasifikasi Suara Paru-Paru Berdasarkan Ciri MFCC

Dody Rafiqo^{*1}, Yohanes Suyanto², Catur Atmaji³

¹Prodi Elektronika dan Instrumentasi FMIPA, UGM Yogyakarta, Indonesia

^{2,3}Departemen Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta, Indonesia
e-mail: ^{*1}dodyrafiqo99@mail.ugm.ac.id, ²yanto@ugm.ac.id, ³catur_atmaji@ugm.ac.id

Abstrak

Paru-paru merupakan organ penting dalam sistem pernapasan manusia, berfungsi untuk menukarkan karbon dioksida dari dalam darah dengan oksigen yang ada di udara. Deteksi adanya gangguan pernapasan dan gangguan pada paru-paru dapat dilakukan melalui berbagai cara; melihat rekam medis, pemeriksaan fisik, pendeteksian dengan x-ray dan juga auskultasi pernapasan. Pemrosesan sinyal digital dapat digunakan sebagai salah satu metode untuk mengetahui adanya gangguan pada paru-paru berdasarkan suara yang dihasilkan. Pada penelitian ini dilakukan pengklasifikasian suara paru-paru pada kelas normal, crackle, wheeze, dan crackle-wheeze dengan menggunakan metode Mel Frequency Cepstral Coefficient (MFCC) dan Convolutional Neural Network (CNN).

Pengamatan dilakukan dengan melakukan variasi pada ekstraksi ciri MFCC dengan menggunakan MFCC 8 dan 13 koefisien, jumlah frame 50 dan 60, dan lebar frame yang digunakan 0,1, 0,15 dan 0,2 detik. Hasil ekstraksi ciri kemudian diterapkan pada sistem klasifikasi CNN, serta confusion matrix digunakan untuk mendapatkan nilai akurasi dan presisi. Nilai akurasi dan presisi tertinggi didapatkan sebesar 71,85% dan 65,70% pada MFCC 13 koefisien dengan rata-rata 71,18%. Berdasarkan hasil tersebut, sistem yang telah dibuat dapat mengklasifikasi suara paru-paru normal, crackle, wheeze dan crackle-wheeze dengan cukup baik.

Kata kunci—suara paru-paru, crackle, wheeze, frame MFCC, CNN

Abstract

The lungs are an important organ in the human respiratory system, which functions to exchange carbon dioxide from the blood with oxygen in the air. Detection of respiratory disorders and lung disorders can be done in various ways; view medical records, physical examination, detection by x-ray and also auscultation of breathing. Digital signal processing can be used as a method to detect lung disorders based on the sound produced. In this study, lung sounds were classified into normal, crackle, wheeze, and crackle-wheeze classes using the Mel Frequency Cepstral Coefficient (MFCC) and Convolutional Neural Network (CNN) methods.

Observations were made by varying the MFCC feature extraction using MFCC 8 and 13 coefficients, the number of frames are 50 and 60, and the width of the frames used was 0,1, 0,15 and 0,2 seconds. The result of feature extraction is then applied to the CNN classification system, and the confusion matrix is used to get the accuracy and precision values. The highest accuracy and precision values were obtained at 71,85% and 65,70% on the MFCC 13 coefficient with an average of 71,18%. Based on these results, the system that has been created can classify normal lung sounds, crackle, wheeze and crackle-wheeze quite well.

Keywords—lung sounds, crackle, wheeze, MFCC's frame, CNN

1. PENDAHULUAN

Paru-paru merupakan organ penting dalam sistem pernapasan manusia, berhubungan dengan sistem ekskresi dan sistem peredaran darah. Paru-paru memiliki fungsi utama yaitu menukarkan karbon dioksida dari dalam darah dengan oksigen yang ada di udara. Pola hidup manusia dan lingkungan sangat berpengaruh terhadap kesehatan organ utama pernapasan. Gangguan pada paru-paru biasanya disebabkan oleh adanya infeksi bakteri, virus atau jamur, dan didukung dengan lingkungan dengan banyak asap kendaraan dan rokok. Apabila terdapat gangguan pada paru-paru manusia maka sistem pernapasan juga ikut terganggu, bahkan resiko terburuknya bisa menyebabkan kematian [1].

Deteksi adanya gangguan pernapasan dan paru-paru dapat dilakukan melalui berbagai cara; melihat rekam medis, pemeriksaan fisik, pendeteksian dengan x-ray dan juga auskultasi pernapasan [2]. Pemrosesan sinyal digital dapat menjadi salah satu metode yang digunakan untuk mengetahui adanya gangguan pada paru-paru berdasarkan suara yang dihasilkan. Pada penelitian sebelumnya dengan menggunakan *dataset ICBHI 2017 Challenge*, metode *low level feature* dan *decision tree* menghasilkan akurasi 49,62% untuk mengklasifikasikan suara paru-paru [3]. Sedangkan penelitian dengan metode ekstraksi ciri STFT dan Wavelet serta SVM sebagai metode klasifikasi menghasilkan akurasi sebesar 57,88% [4].

Pada penelitian yang berhubungan dengan *speech processing* dan *speech recognition*, *Mel Frequency Cepstral Coefficient* (MFCC) adalah salah satu metode fitur ekstraksi yang sering digunakan. MFCC memiliki beberapa keunggulan yaitu bisa menangkap informasi-informasi penting yang terdapat dalam data suara. MFCC mampu menghasilkan data dengan jumlah minimal tanpa mengurangi informasi-informasi penting yang ada. MFCC mengadaptasi telinga manusia dalam melakukan persepsi terhadap data suara. Fungsi logaritmik pada MFCC mampu menghasilkan pendekatan yang lebih baik terhadap pendengaran manusia [5].

Convolutional Neural Network (CNN) merupakan salah satu jenis *neural network* yang sering digunakan dalam proses *learning* dalam suatu sistem. CNN terdiri dari beberapa *layer* dalam melakukan klasifikasi meliputi lapisan *convolution*, lapisan *pooling* dan lapisan *fully connected*. Dengan adanya proses konvolusi, data dengan struktur dua dimensi seperti data gambar akan sangat cocok jika diproses menggunakan CNN [6].

Berdasarkan hal diatas, Peneliti tertarik untuk dalam melakukan klasifikasi suara paru-paru jenis normal dan suara paru-paru yang memiliki gangguan (abnormal). MFCC dapat berfungsi sebagai vektor ciri yang baik untuk merepresentasikan suara manusia dan suara musik. CNN digunakan sebagai algoritma dalam pembuatan model klasifikasi. Pada penelitian ini suara akan dikelompokkan menjadi empat kelas yaitu suara normal, suara *crackles*, suara *wheeze* dan suara *crackle-wheeze*.

2. METODE PENELITIAN

Pada penelitian ini akan dibuat sistem yang mampu mengklasifikasikan suara paru-paru manusia menjadi empat jenis kelompok yaitu suara normal, *crackle*, *wheeze* dan *crackle-wheeze*. *Dataset* suara yang digunakan sebagai masukan pada penelitian ini menggunakan data dari ICBHI 2017 *Challenge*. Suara dari keempat kelas selanjutnya akan masuk kedalam proses *pre-processing* yakni dengan penyamaan *sampling rate*, normalisasi dan *butterworth filter*, kemudian ekstraksi ciri dengan MFCC dan klasifikasi dengan CNN. Diagram blok sistem ditunjukkan oleh Gambar 1.



Gambar 1 Diagram blok sistem

2.1 Dataset

Dataset diperoleh dari ICBHI 2017 *Challenge database* yang dapat diakses melalui Internet. Pengambilan data suara paru-paru dilakukan pada beberapa bagian tubuh pasien; *trachea* (Tc), *anterior left* (Al), *anterior right* (Ar), *posterior left* (Pl), *posterior right* (Pr), *lateral left* (Ll), and *lateral right* (Lr). Database memiliki total durasi 5.5 jam rekaman yang terdiri dari 6898 siklus pernapasan, 1864 di antaranya berisi *crackles*, 886 berisi *wheeze*, 3642 berisi normal, 506 berisi *crackle-wheeze* dari 126 subjek [7]. Data diambil dengan menggunakan 3 instrumen yang berbeda dengan *sampling rate* yaitu 4 kHz, 10 kHz, dan 44,1 kHz.

2.2 Pre-processing

Pre-processing dilakukan untuk menyiapkan data yang akan diolah ke tahap selanjutnya. Terdapat dua tahap proses *pre-processing* yakni penyamaan *sampling rate*, normalisasi dan *butterworth filter* yang masing-masing digunakan untuk menyeragamkan baik dari segi jangkauan data maupun rentang frekuensinya.

2. 2.1 Up/Down sampling

Dataset yang digunakan pada penelitian ini memiliki *sampling rate* yang berbeda yaitu 4 kHz, 10 kHz, dan 44,1 kHz. Untuk memudahkan proses pengolahan pada sinyal dilakukan *up/down sampling* menjadi 11025 Hz.

2. 2.2 Normalisasi

Normalisasi digunakan untuk menyeragamkan jangkauan data yakni diantara -1 sampai 1. Hal ini diharapkan bisa untuk menyeragamkan data, tidak ada data yang terlalu tinggi dan tidak ada yang terlalu rendah. Cara yang dilakukan untuk mendapatkan hasil tersebut adalah dengan membagi dengan data maksimum di setiap data sinyal suara atau bisa dihitung dengan Persamaan (1)

$$\|x(n)\| = \frac{x(n)}{\max(\text{abs}(x(n)))} \quad (1)$$

keterangan:

$\|x(n)\|$: sinyal hasil normalisasi

$x(n)$: sinyal setelah *pre-emphasize*

2.2.3 Butterworth Filter

Butterworth filter merupakan salah satu jenis *filter* yang penggunaannya dapat menghasilkan frekuensi respon yang sedatar mungkin pada *passband*. Diharapkan dengan *butterworth filter* dapat membuat data-data yang akan dilakukan ekstraksi ciri memiliki rentang frekuensi yang sama sehingga akan memudahkan proses selanjutnya. Spesifikasi dari *butterworth filter* yang digunakan adalah *butterworth filter* orde 5 dengan rentang frekuensi antara 100 hingga 4000 Hz. Rentang frekuensi tersebut digunakan untuk menghilangkan artefak yang disebabkan oleh detak jantung, ambien dan gerakan tubuh [8]. Dengan menggunakan filter ini frekuensi sinyal yang lebih rendah dari 100 Hz dan lebih tinggi dari 4000 Hz akan ditahan.

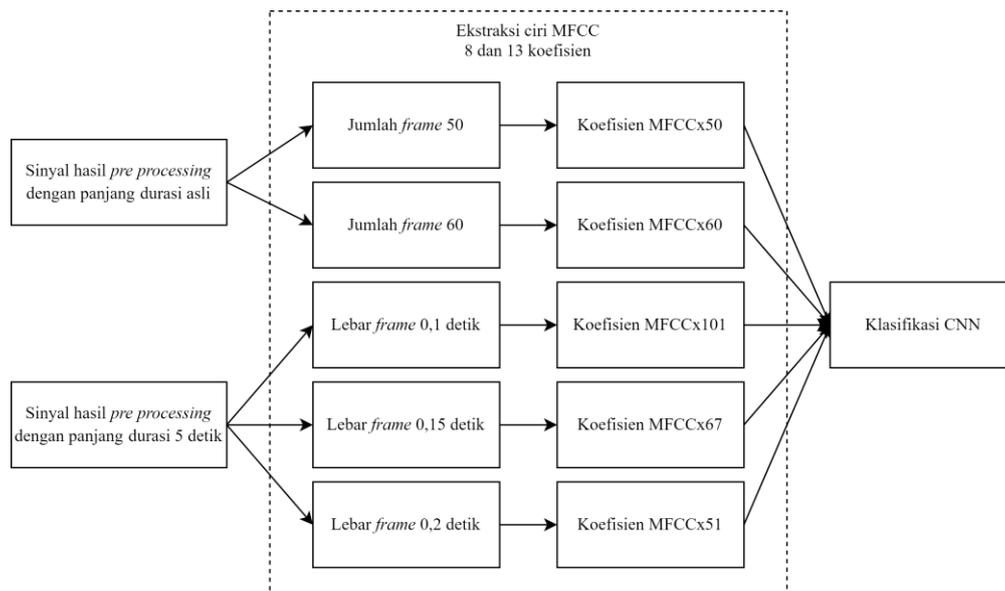
2. 3 MFCC

Tahap fitur ekstraksi dilakukan menggunakan metode *Mel Frequency Cepstral Coefficient* (MFCC). MFCC memiliki kemampuan dominan di bidang pengenalan suara karena dapat merepresentasikan spektrum amplitudo suara dalam bentuk yang kompak dan tetap mempertimbangkan persepsi sistem pendengaran manusia. Proses awal MFCC dimulai dengan mengaplikasikan *pre-emphasis* dengan koefisien α yang biasa digunakan sebesar 0,97. Selanjutnya sinyal suara dibagi kedalam *frame-frame* kecil dan terjadi proses *overlapping*

sebesar 50%. Pada setiap *frame* yang ada diaplikasikan jendela *hamming* dalam fungsi *windowing*. Pada proses pemotongan sinyal dimungkinkan terjadi ketidak kontinyuan pada awal dan akhir *frame*, sehingga diperlukan fungsi *windowing* untuk mengatasi aliasing pada sinyal. Kemudian dilakukan perubahan sinyal dari domain waktu ke dalam domain frekuensi dengan menggunakan *Fast Fourier Transform* dengan besar nilai analisa yaitu 2^n . Tahap selanjutnya dari MFCC yaitu dilakukan penapisan menggunakan *filterbank* pada setiap sinyal yang berada dalam domain frekuensi menjadi *mel scale*. Tahap akhir dari MFCC yaitu pengaplikasian logaritma cepstrum dan DCT untuk mengubah sinyal dari ranah frekuensi ke ranah waktu. Proses MFCC menghasilkan fitur ekstraksi dalam bentuk matriks. Besar dimensi matriks MFCC dipengaruhi oleh nilai koefisien MFCC dan jumlah *frame* yang digunakan. Maka dari itu, dimensi matriks ciri MFCC akan memiliki ukuran yang berbeda pada data yang memiliki berdurasi waktu yang berbeda [9].

Pada penelitian ini ditentukan koefisien MFCC dan jumlah *frame* yang akan digunakan yaitu 8 dan 13 untuk koefisien MFCC, 50 dan 60 untuk jumlah *frame*. Hal ini dilakukan karena sinyal yang digunakan memiliki panjang durasi yang berbeda-beda sehingga apabila jumlah *frame* tidak disamakan maka MFCC akan menghasilkan matriks ciri yang nilainya beragam sehingga sulit digunakan sebagai masukan dari CNN.

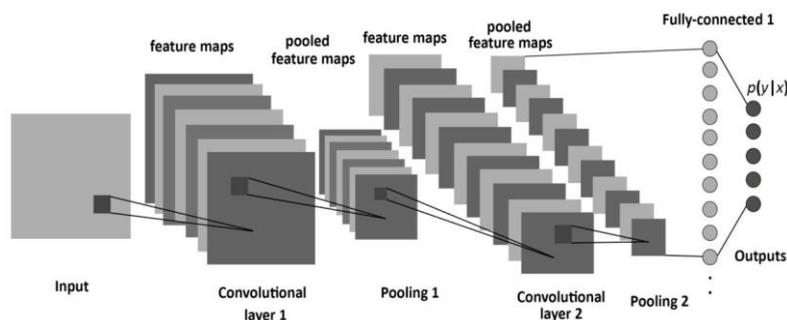
Ekstraksi ciri menggunakan MFCC juga dilakukan dengan melakukan variasi pada lebar *frame* yang digunakan yaitu 0,1, 0,15 dan 0,2 detik pada sinyal yang disamakan durasinya menjadi 5 detik. Matriks hasil MFCC dari satu sinyal akan digabungkan dengan sinyal yang lain untuk masuk ke tahap klasifikasi. Proses ekstraksi ciri MFCC ditunjukkan pada Gambar 2.



Gambar 2 Variasi parameter pada proses ekstraksi ciri

2. 4 Klasifikasi

Convolutional Neural Network (CNN) merupakan salah satu jenis *neural network* yang sering digunakan untuk memproses data citra. Meskipun demikian, CNN dapat digunakan pada data suara dengan memodifikasi sinyal inputan menjadi bentuk 2 dimensi sehingga dapat digunakan sebagai masukan CNN. Parameter yang digunakan CNN disebut dengan kernel dan dapat dianggap sebagai detektor untuk pola lokal pada data. Secara garis besar CNN tersusun oleh neuron yang memiliki *weight*, bias dan fungsi aktivasi. CNN terdiri dari beberapa lapisan dalam melakukan klasifikasi meliputi lapisan *convolution*, lapisan *pooling* dan lapisan *fully connected* [6]. Arsitektur CNN secara umum ditunjukkan pada Gambar 3.



Gambar 3 Arsitektur CNN

Convolutional layer adalah lapisan pertama yang menerima input langsung pada arsitektur dan menjadi pondasi dari arsitektur CNN. Pada lapisan ini terjadi operasi konvolusi yaitu operasi kombinasi *linier filter* terhadap daerah lokal. *Convolutional layer* melakukan operasi konvolusi terhadap keluaran dari *layer* sebelumnya. Operasi konvolusi terhadap data masukan bertujuan untuk mendapatkan fitur ekstraksi ciri [10].

Pooling layer berfungsi untuk *downsampling output* dari *convolutional layer* sepanjang dimensi spasial tinggi dan lebar. Data dipecah menjadi beberapa bagian berdasarkan ukuran *layer* yang sudah ditetapkan. *Max pooling* menjadi salah satu metode yang umum digunakan pada lapisan *pooling*, yaitu dengan memilih nilai paling besar pada matriks data [11].

Lapisan *fully-connected* sama dengan jaringan syaraf tiruan yang umum digunakan dalam proses *learning*, yaitu semua neuron aktivasi dari lapisan sebelumnya terhubung dengan neuron di lapisan selanjutnya. Setiap aktivasi dari lapisan sebelumnya perlu diubah menjadi data satu dimensi sebelum dapat dihubungkan ke semua neuron di lapisan [12]. Lapisan *fully-connected* memiliki tujuan untuk memproses data sehingga mudah diklasifikasikan, lapisan ini umumnya digunakan pada *multi layer perceptron*. Lapisan *fully-connected* memiliki perbedaan yang mendasar apabila dibandingkan dengan lapisan konvolusi yaitu terdapat pada neuron yang terhubung secara keseluruhan, pada lapisan konvolusi neuron hanya terhubung pada daerah tertentu saja.

Pada proses *neural network* terdapat fungsi aktivasi. Fungsi aktivasi ini hanya bekerja pada neuron lapisan tersembunyi dan lapisan *output*, tidak bekerja pada lapisan *input*. *Rectified Linear Unit* (ReLU) merupakan salah satu jenis fungsi aktivasi yang sering digunakan. ReLU bekerja dengan cara Menghasilkan nilai 0 pada data yang bernilai negatif.

Arsitektur CNN yang digunakan terdiri atas 3 buah *convolutional layer* dan 3 buah *fully-connected layer*. Pada setiap *convolutional layer* digunakan ukuran *kernel* 3×3 , *stride* 1, dan *max pooling* 2×2 . Filter yang digunakan akan meningkat pada setiap lapisan mulai dari 16, 32 dan 64. Lapisan pertama akan menerima input berupa matriks hasil ekstraksi ciri (koefisien MFCC, jumlah *frame*). Pada *fully-connected layer* digunakan beberapa jumlah neuron yaitu 2048, 1024 dan 128. ReLU digunakan sebagai fungsi aktivasi pada lapisan ini. Lapisan keluaran dari CNN terdiri dari 4 node yang menggunakan fungsi aktivasi *softmax*.

2.5 Pengujian

Pengujian dilakukan dengan menggunakan *k fold cross validation*. *K fold cross validation* adalah salah satu metode yang berfungsi untuk mengevaluasi sebuah model dengan memisahkan dataset menjadi data latih dan data uji sehingga model akan diimplementasikan dengan iterasi sebanyak jumlah *fold* dengan kombinasi data pengujian dan data pelatihan berdasarkan kombinasi disetiap *fold* [13]. Jumlah nilai *k* yang digunakan adalah 5 agar data pelatihan dan data pengujian memiliki jumlah yang cukup. Artinya jumlah keseluruhan data akan dibagi menjadi 5 bagian dengan satu bagian menjadi data pengujian dan 4 bagian lain menjadi data pelatihan. Sistem diimplementasikan dengan dataset tersebut sehingga sistem dijalankan sebanyak 5 iterasi dengan pembagian data uji sebanyak 1380 data

(20%) dan data latih sebanyak 5518 data (80%). Dilakukan perulangan dengan mengubah bagian yang sebelumnya menjadi data latih menjadi data uji hingga keseluruhan data mendapatkan giliran untuk menjadi data uji atau seperti yang terlihat pada Tabel 1.

Tabel 1 5 fold cross validation

Fold ke-	Bagian data ke-				
	1	2	3	4	5
Fold 1					
Fold 2					
Fold 3					
Fold 4					
Fold 5					

	Data validasi
	Data latih

Setelah mendapatkan hasil pelatihan dan pengujian dari masing-masing *fold*, didapatkan *confusion matrix*. *Confusion matrix* berfungsi untuk merepresentasikan kerja sistem dengan cara membandingkan prediksi dari sistem dengan kelas yang sebenarnya [14]. Hal itu bisa digunakan untuk mendapatkan nilai dari *True Positif* (TP), *False Positif* (FP), *False Negatif* (FN), dan *True Negatif* (TN). Sehingga bisa ditentukan nilai akurasi, presisi, recall/TPR, dan FDR dengan Persamaan (2),(3),(4), dan (5).

$$akurasi = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% \quad (2)$$

$$Presisi = \frac{\sum_{i=1}^l FP_i}{\sum_{i=1}^l TP_i + FP_i} \times 100\% \quad (3)$$

$$recall = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FN_i} \times 100\% \quad (4)$$

$$FDR = \frac{\sum_{i=1}^l FP_i}{\sum_{i=1}^l FP_i + TP_i} \quad (5)$$

Dengan keterangan :

l : banyak kelas

TP_i : banyak data positif yang diprediksi benar masuk untuk kelas ke= i

TN_i : banyak data positif yang diprediksi salah masuk untuk kelas ke= i

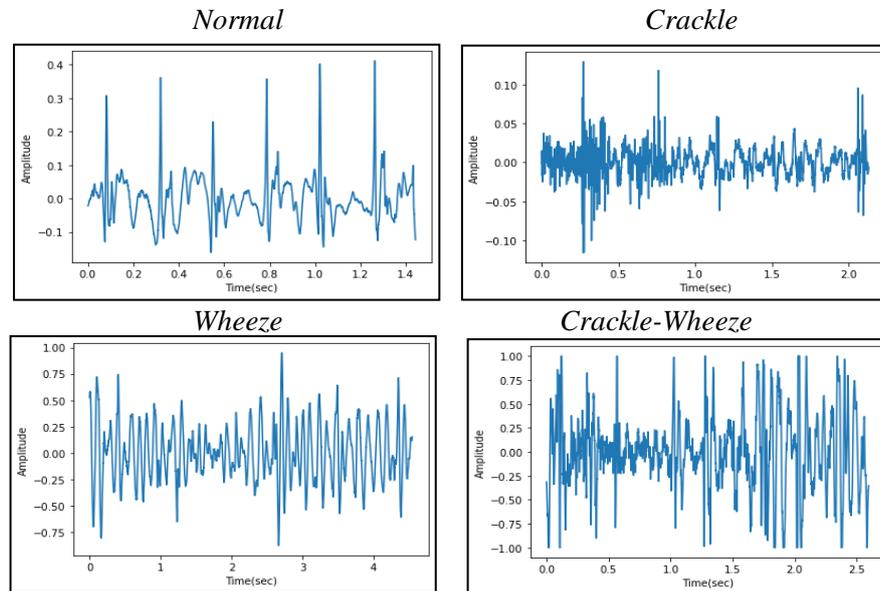
FP_i : banyak data negatif yang diprediksi benar masuk untuk kelas ke= i

FN_i : banyak data negatif yang diprediksi salah masuk untuk kelas ke= i

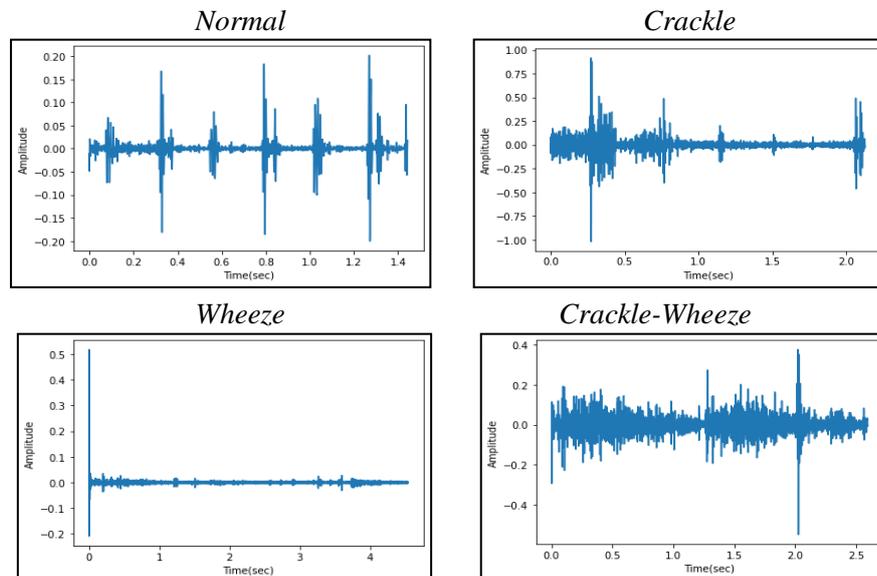
3. HASIL DAN PEMBAHASAN

Data suara paru-paru terdiri dari 6898 siklus pernapasan, 1864 di antaranya berisi *crackles*, 886 berisi *wheeze*, 3642 berisi normal, 506 berisi *crackle-wheeze* dari 126 subjek. Sinyal suara yang menjadi masukan sistem berupa disamakan nilai frekuensi samplingnya menjadi 11025. Sinyal kemudian masuk ke tahap *pre processing* untuk menyamakan rentang nilai dengan normalisasi dan menyamakan rentang frekuensi dengan *butterworth filter*.

Butterworth filter yang digunakan adalah orde 5 dengan frekuensi sekitar 100-4000 Hz, untuk mengurangi *noise* pada *rentang* frekuensi tersebut sekaligus tetap menjaga ciri dari sinyal. Sinyal sebelum dan sesudah *pre processing* ditunjukkan pada Gambar 4 dan Gambar 5.



Gambar 4 Sinyal suara paru-paru



Gambar 5 Sinyal setelah dikenai *butterworth filter*

Data yang telah melalui *pre processing* kemudian akan diekstrak cirinya dengan metode MFCC. MFCC dilakukan dengan menggunakan *library* librosa dengan mengatur nilai *overlapping*, *sampling rate*, jumlah *frame* pada sinyal dan koefisien MFCC yang digunakan. *Overlapping* yang digunakan pada penelitian ini yaitu 50%. Pengaturan pada jumlah *frame* sinyal mengakibatkan lebar *frame* dari masing-masing sinyal memiliki panjang yang berbeda-beda. Hal tersebut dapat terlihat pada Tabel 2 dengan menggunakan 4 contoh sinyal yang berbeda.

Tabel 2 Pengaruh jumlah *frame* terhadap lebar *frame* sinyal

Durasi asli sinyal dalam detik	Banyak data dalam 1 <i>frame</i>	
	<i>frame</i> 50	<i>frame</i> 60
2,128	957	795
2,596	1168	970
1,443	649	539
4,543	2044	1697

Hasil dari proses ekstraksi ciri akan menghasilkan matriks ciri dengan ukuran koefisien MFCC×jumlah *frame*. Pada penelitian ini digunakan koefisien MFCC bernilai 8 dan 13, sedangkan jumlah *frame* yang digunakan adalah 50 dan 60 *frame*. Sehingga didapatkan 4 ukuran matriks yang berbeda yaitu 8×50, 8×60, 13×50, dan 13×60. Data tersebut kemudian dikumpulkan bersama dengan data yang berasal dari *file* lain sehingga untuk keseluruhan *dataset* akan membentuk matriks 6898×8×50, 6898×8×60, 6898×13×50, dan 6898×13×60 dimana 6898 merupakan jumlah total data yang digunakan.

Selain variasi pada jumlah *frame*, ekstraksi ciri menggunakan MFCC dapat dilakukan dengan mengubah nilai lebar *frame*. Sinyal yang sudah disamakan durasinya menjadi 5 detik akan dilakukan proses MFCC dengan memvariasikan lebar *frame*. Pada penelitian ini digunakan tiga nilai lebar *frame* yaitu 0.1 detik (1102 data), 0.15 detik (1653 data) dan 0.2 detik (2205 data) pada koefisien MFCC 8 dan 13. Variasi nilai lebar *frame* mengakibatkan jumlah *frame* yang dihasilkan MFCC bervariasi seperti yang terlihat pada Tabel 3.

Tabel 3 Pengaruh lebar *frame* terhadap jumlah *frame*

Lebar <i>Frame</i>	Jumlah <i>Frame</i>
0,1 detik	101 <i>frame</i>
0,15 detik	67 <i>frame</i>
0,2 detik	51 <i>frame</i>

Hasil MFCC dengan variasi lebar *frame* tersebut menghasilkan matriks ciri dengan ukuran 8×101, 8×67, 8×51 dan 13×101, 13×67, 13×51. Kemudian data tersebut akan digabungkan dengan data yang lain sehingga menghasilkan matriks berukuran 6898×8×101, 6898×8×67, 6898×8×51 dan 6898×13×101, 6898×13×67, 6898×13×51.

Hasil dari MFCC tersebut menjadi ciri yang akan masuk ke dalam klasifikasi menggunakan CNN. Hasil dari proses klasifikasi berupa persentase sinyal tersebut masuk ke dalam sebuah kelas. Kelas prediksi dari sinyal tersebut merupakan kelas dengan jumlah prediksi terbesar daripada kelas yang lainnya. Setelah dilakukan pengujian dengan menggunakan 4 fold *cross validation*, didapatkan hasil berupa *confusion matrix*. Berdasarkan tabel *confusion matrix*, bisa dihitung TP, FN, FN, TN dan nilai akurasi, presisi, TPR, FDR dari masing-masing pengujian seperti pada Tabel 4 dan Tabel 5.

Tabel 4 Rata-rata akurasi, TPR, FDR dan presisi pada sistem MFCC 8 koefisien

Parameter	Pelatihan				Pengujian			
	Akurasi (%)	TPR (%)	FDR (%)	Presisi (%)	Akurasi (%)	TPR (%)	FDR (%)	Presisi (%)
nframe 50	89,33	83,68	10,05	89,95	71,41	59,31	31,92	68,08
nframe 60	86,71	80,74	12,17	87,83	69,96	59,82	33,23	66,71
lframe 0.1 detik	85,53	79,46	14,94	85,06	69,37	58,72	36,01	63,99
lframe 0.15 detik	84,11	74,61	14,83	85,17	70,5	57,18	33,2	66,8
lframe 0.2 detik	83,73	74,64	14,68	85,32	69,73	56,68	35,77	64,23
Rata-rata	85,88	78,63	13,33	86,67	70,19	58,34	34,03	65,96
Keterangan: nframe: jumlah <i>frame</i> yang digunakan tetap, lebar <i>frame</i> yang digunakan menyesuaikan lframe: lebar <i>frame</i> yang digunakan tetap, jumlah <i>frame</i> yang dihasilkan menyesuaikan								

Tabel 5 Rata-rata akurasi, TPR, FDR dan presisi pada sistem MFCC 13 koefisien

Parameter	Pelatihan				Pengujian			
	Akurasi (%)	TPR (%)	FDR (%)	Presisi (%)	Akurasi (%)	TPR (%)	FDR (%)	Presisi (%)
nframe 50	91,39	88,47	8,91	91,09	71,48	61,09	34,03	65,97
nframe 60	91,42	88,53	8,56	91,44	70,76	60,11	35,01	64,99
lframe 0.1	88,72	84,39	12,1	87,9	71,85	61,04	34,3	65,7
lframe 0.15 detik	85,15	78,02	14,1	85,9	70,32	58,18	34,29	65,71
lframe 0.2 detik	88,06	81,77	11,74	88,26	71,5	59,75	34,16	65,84
Rata-rata	88,95	84,24	11,08	88,92	71,18	60,03	34,36	65,64
Keterangan: nframe: jumlah <i>frame</i> yang digunakan tetap, lebar <i>frame</i> yang digunakan menyesuaikan lframe: lebar <i>frame</i> yang digunakan tetap, jumlah <i>frame</i> yang dihasilkan menyesuaikan								

Pada sistem dengan menggunakan MFCC 8 koefisien memiliki rata-rata nilai akurasi, TPR, FDR dan presisi yaitu sebesar 85,88%, 78,62%, 13,33% dan 86,67% pada proses pelatihan dan 70,19%, 58,34%, 34,02% dan 65,96% pada proses pengujian. Akurasi tertinggi didapatkan pada sistem menggunakan *frame* berjumlah 50 yaitu 71,41%. Akurasi terendah didapatkan pada sistem menggunakan lebar *frame* 0,1 detik yaitu 69,37%.

Pada sistem dengan menggunakan MFCC 13 koefisien memiliki rata-rata nilai akurasi, TPR, FDR dan presisi yang lebih tinggi daripada menggunakan MFCC 8 koefisien yaitu sebesar 88,95%, 84,24%, 11,08% dan 88,92% pada proses pelatihan dan 71,18%, 60,03%, 34,36% dan 65,64% pada proses pengujian. Akurasi tertinggi didapatkan pada sistem menggunakan lebar *frame* 0,1 detik yaitu 71,85%. Akurasi terendah didapatkan pada sistem menggunakan lebar *frame* 0,15 detik yaitu 70,32%.

Sistem dengan koefisien MFCC yang sama ketika dilakukan pengujian menggunakan variasi jumlah *frame* dan lebar *frame* memiliki nilai akurasi cenderung sama. Sedangkan, ketika menggunakan koefisien MFCC yang berbeda dengan jumlah *frame* dan lebar *frame* yang sama memiliki perbedaan akurasi yang cukup terlihat. Hal ini menunjukkan bahwa variasi pada koefisien MFCC lebih berpengaruh terhadap akurasi sistem.

Penggunaan suara paru-paru dengan durasi asli sinyal yang berbeda-beda dibandingkan dengan sinyal yang disamakan durasinya menjadi 5 detik menggunakan *zero padding* memiliki akurasi sistem yang cenderung sama. Hal ini menunjukkan penambahan *zero padding* pada sinyal tidak menurunkan kualitas sistem yang dibangun.

Hasil dari penelitian ini bisa dibandingkan dengan penelitian yang telah dilakukan sebelumnya ditunjukkan pada Tabel 6.

Tabel 6 Pembeding penelitian sebelumnya

Peneliti	Objek Klasifikasi	Metode ekstraksi ciri dan klasifikasi	akurasi
Chambres <i>et al.</i> , 2018	Suara paru-paru normal dan tidak normal (<i>crackle</i> , <i>wheeze</i> dan keduanya)	<i>Low level feature</i> dan <i>Decision Tree</i>	50%
Serbes <i>et al.</i> , 2018	Suara paru-paru normal, <i>crackle</i> , <i>wheeze</i> dan keduanya	STFT+Wavelet dan SVM	57,88%
Hasanain <i>et al.</i> , 2018	Suara paru-paru normal, <i>crackle</i> , <i>wheeze</i> dan keduanya	<i>Spectrogram</i> dan CNN	74%
Rizal <i>et al.</i> , 2018	Suara bronchial, broncovesikular, tracheal dan vesikular	LPC dan JST <i>Backpropagation</i>	98,33%

Nilai akurasi dari penelitian ini memang tidak terlalu bagus apabila dibandingkan dengan hasil penelitian yang dilakukan oleh Rizal [15]. Hal tersebut dikarenakan selain menggunakan metode yang berbeda, kelas penelitian yang diklasifikasikan juga berbeda, penelitian ini berfokus pada klasifikasi suara paru menjadi kelas normal, *crackle*, *wheeze* dan *crackle-wheeze*, sedangkan penelitian yang dilakukan oleh Rizal dkk berfokus pada klasifikasi suara paru-paru normal menjadi kelas *broncial*, *broncovesikular*, *tracheal* dan *vesikular* [15].

Penelitian ini memperoleh nilai akurasi yang lebih baik dibandingkan dengan penelitian sebelumnya dengan menggunakan *dataset* yang sama dalam mengelompokkan suara paru-paru kedalam 4 kelas yaitu kelas normal, *crackle*, *wheeze* dan *crackle-wheeze*. Penelitian sebelumnya yang dilakukan oleh Serbes dkk memiliki nilai akurasi tertinggi sebesar 57,88% [4], sedangkan penelitian ini dengan menggunakan metode MFCC untuk ekstraksi ciri dan CNN untuk proses klasifikasi, mendapatkan akurasi tertinggi yaitu 71,85%.

Penelitian yang dilakukan oleh Hasanain dkk memiliki akurasi 74%, nilai tersebut tidak terlalu berbeda jika dibandingkan dengan akurasi penelitian ini, yaitu 71,85%. Dengan menggunakan metode ekstraksi ciri yang berbeda dan jenis klasifikasi yang sama, penelitian ini memiliki selisih akurasi sebesar 2,15%. Penelitian yang dilakukan oleh Hasanain dkk menggunakan *scale* yang berkisar 1-8 dan ukuran dalam proses *spectrogram* dari sinyal sebesar 1024 yang disesuaikan dengan standar ukuran citra sehingga cocok dikombinasikan dengan CNN yang sudah dikenal sebagai pengklasifikasi objek berbasis citra [2].

Perbedaan penelitian ini dibandingkan penelitian sebelumnya dengan menggunakan data yang sama adalah karena penelitian ini menggunakan metode MFCC pada tahapan ekstraksi ciri. Proses MFCC mudah dilakukan dan dapat dilakukan banyak variasi pada parameter-parameter yang digunakan seperti koefisien MFCC, jumlah *frame*, lebar *frame* dan nilai *overlapping* sehingga didapatkan ekstraksi ciri yang baik untuk proses klasifikasi.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, yaitu klasifikasi suara paru-paru menggunakan *Mel Frequency Cepstral Coefficien* (MFCC) dapat disimpulkan bahwa pengujian menggunakan MFCC sebagai ekstraksi ciri dan *Convolutional Neural Network* (CNN) sebagai pengklasifikasi memberikan hasil akurasi rata-rata sebesar 71,18% pada MFCC 13 koefisien, dengan akurasi terbaik yaitu 71,85%. Dengan menggunakan *dataset* yang sama, sistem mendapatkan nilai akurasi yang lebih tinggi dibandingkan dengan nilai akurasi pada penelitian sebelumnya yang menggunakan STFT dan *Wavelet* sebagai metode ekstraksi ciri dan SVM untuk metode klasifikasi, dengan akurasi terbaik sebesar 57,88%. Hal ini menunjukkan bahwa metode MFCC dan CNN terbukti cukup baik untuk digunakan sebagai metode pengolahan data yang diteliti.

DAFTAR PUSTAKA

- [1] F. Syafria, A. Buono, and B. P. Silalahi, "Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier," *J. Ilmu Komput. dan Agri-Informatika*, vol. 3, no. 1, p. 27, 2017.
- [2] I. W. Hasanain, A. Rizal, and Jondri, "Klasifikasi Suara Paru-Paru Menggunakan Convolutional Neural Network (CNN)," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3218–3223, 2021.

- [3] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," *Proc. - Int. Work. Content-Based Multimed. Index.*, 2018.
- [4] G. Serbes, S. Ulukaya, and Y. P. Kahya, "An automated lung sound preprocessing and classification system based on spectral analysis methods," *IFMBE Proceedings*, vol. 66, pp. 45–49, 2018.
- [5] M. A. Romli and A. Solichin, "Pemrosesan Sinyal Digital Untuk Mengidentifikasi Akord Dasar Penyanyi Dengan Metode Mel Frequency Cepstral Coefficients (MFCC) Dan Jaringan Syaraf Tiruan Backpropagation. Digital Signal," *Semin. Nas. Multidisiplin Ilmu*, 2017.
- [6] I. B. L. M. Suta, R. S. Hartati, and Y. Divayana, "Diagnosa Tumor Otak Berdasarkan Citra MRI (Magnetic Resonance Imaging).," *JTE 18*, 149–154, 2019.
- [7] B. M. Rocha *et al.*, "A respiratory sound database for the development of automated classification," *IFMBE Proc.*, vol. 66, pp. 33–37, 2018.
- [8] P. S. Faustino, "Crackle and wheeze detection in lung sound signals using convolutional neural networks," 2019. [online]. Available: <https://repositorio-aberto.up.pt/bitstream/10216/125107/2/372897>. [Accessed: 20-Okt-2020]
- [9] J. K. A. Christya, C. Atmaji, and A. E. Putra, "Deteksi Kesalahan Pengucapan Huruf Jawa Carakan dengan Jaringan Syaraf Tiruan Perambatan Balik," *IJEIS*, pp. 1–12, 2021.
- [10] S. R. Dewi, "Deep Learning Object Detection pada Video menggunakan Tensorflow dan Convolutional Neural Network 95," *Skripsi*, FMIPA, Universitas Islam Indonesia, Yogyakarta, 2018.
- [11] R. A. Pangestu, B. Rahmat, and F. T. Anggraeny, "Implementasi Algoritma CNN untuk Klasifikasi Citra Lahan dan Perhitungan Luas," *Inform. dan Sist. Inf.*, vol. 1, no. 1, pp. 166–174, 2020.
- [12] F. Prayogi, "Identifikasi Lahan Pertanian Menggunakan Convolutional Neural Network (CNN) Pada Citra Google Earth 188," 2020. [online]. Available: https://github.com/SoedirmanMachineLearning/Identification_paddy_field. [Accessed: 25-Oct-2020]
- [13] A. Wibowo, "10 Fold-Cross Validation," 2017. [online]. available: <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>. [Accessed: 23-Sept-2021]
- [14] S. Visa, B. Ramsay, A. Ralescu, and E. van der Knaap, "Confusion Matrix-based Feature Selection," *Conference Paper*, vol. 710, pp.1-8.2011.

- [15] A. Rizal, L. Anggraeni, and V. Suryani, "Pengenalan Suara Paru-Paru Normal Menggunakan LPC dan Jaringan Syaraf Tiruan Back-Propagation," Preceeding Int. Semin. Electr. Power, Electron. Commun. Control. Informatics (EECCIS 2006), pp. 6–10, 2006.