

# Metode Wavelet-MFCC dan Korelasi dalam Pengenalan Suara Digit

*(Wavelet-MFCC Method and Correlation in Digit Voice Recognition)*

Zaurarista Dyarbirru<sup>[1]\*</sup>, Syahroni Hidayat<sup>[2]</sup>

<sup>[1]</sup>Program Studi Ilmu Komputer, Universitas Bumigora

<sup>[2]</sup>Lembaga Pengembangan Ilmu Pengetahuan dan Teknologi, Sekawan Institute

E-mail: [zaurarista17@gmail.com](mailto:zaurarista17@gmail.com), [s.hidayat@universitasbumigora.ac.id](mailto:s.hidayat@universitasbumigora.ac.id)

## KEYWORDS:

Automatic Speech Recognition, MFCC method, wavelet, wavelet-MFCC, K-Fold Cross Validation

## ABSTRACT

*Voice is the sound emitted from living things. With the development of Automatic Speech Recognition (ASR) technology, voice can be used to make it easier for humans to do something, for example in the process of typing short messages with voice input. ASR systems generally have 3 main processes, namely pre-processing, feature extraction, and recognition. In ASR, the feature extraction process has a very important role because it can affect the recognition results. The feature extraction methods that are commonly applied to ASR are MFCC and Wavelet. Each of them has advantages and disadvantages. MFCC can work to imitate the way the ear works, but the application of Fourier transforms in the conversion process is prone to loss of information, while Wavelets can map signals into the time and frequency domains at the same time without losing information. Therefore, this study will combine the wavelet feature extraction method and MFCC to maximize the advantages of both. The proposed method is called Wavelet-MFCC. For the recognition process using the correlation method. Determination of system performance using the Word Recognition Rate (WRR) method which is validated with K-Fold Cross Validation with the number of folds is 5. The research dataset used is a secondary dataset containing voice recording digits 0-9 in English. The results show that the digit recognition system that has been built provides the highest average accuracy of 63% for digit 4 (four) by using wavelet daubechies db3 and wavelet dyadic transform method. As for the wavelet decomposition method used, it is concluded that the use of dyadic wavelet transformation is better than wavelet packet.*

## KATA KUNCI:

Sistem pengenalan suara, metode MFCC, wavelet, wavelet-MFCC, K-Fold Cross Validation

## ABSTRAK

*Suara adalah bunyi yang dikeluarkan dari makhluk hidup. Dengan adanya perkembangan teknologi Automatic Speech Recognition (ASR), suara dapat digunakan untuk mempermudah manusia dalam melakukan sesuatu, misalkan pada proses pengetikan pesan singkat dengan input suara. Sistem ASR umumnya memiliki 3 proses utama, yaitu pemrosesan awal, ekstraksi ciri, dan pengenalan. Pada ASR proses ekstraksi ciri memiliki peran yang sangat penting karena dapat mempengaruhi hasil pengenalan. Metode ekstraksi ciri yang sudah umum diterapkan pada ASR adalah MFCC dan Wavelet. Masing-masing memiliki kelebihan dan kekurangan. MFCC dapat bekerja meniru cara kerja telinga namun penerapan transformasi fourier dalam proses konversinya mengakibatkan rentan terjadi kehilangan informasi, sementara Wavelet dapat memetakan sinyal ke dalam domain waktu dan frekuensi dalam waktu bersamaan tanpa kehilangan informasi. Oleh karena itu, penelitian ini akan menggabungkan metode ekstraksi ciri wavelet dan MFCC untuk memaksimalkan kelebihan yang ada pada keduanya. Metode yang diusulkan disebut Wavelet-MFCC. Untuk proses pengenalan suara menggunakan metode korelasi. Penentuan performa sistem menggunakan metode Word Recognition Rate (WRR) yang divalidasi dengan K-Fold Cross Validation dengan jumlah fold adalah 5. Dataset penelitian yang digunakan berupa dataset sekunder berisi suara perekaman digit 0-9 dalam bahasa inggris. Hasil menunjukkan bahwa sistem pengenalan suara digit yang telah dibangun memberikan akurasi rata-rata tertinggi 63% untuk digit 4 (empat) dengan menggunakan wavelet daubechies db3 dan metode transformasi wavelet dyadic. Sedangkan untuk metode dekomposisi wavelet yang digunakan, disimpulkan bahwa penggunaan transformasi wavelet dyadic lebih baik dari wavelet paket.*

## I. PENDAHULUAN

Suara adalah bunyi yang dikeluarkan dari makhluk hidup. Dengan adanya perkembangan teknologi, suara dapat digunakan untuk mempermudah manusia dalam melakukan sesuatu. Selama lebih dari lima dekade, pengenalan suara otomatis telah menjadi bidang penelitian aktif. Pengenalan suara memiliki potensi besar untuk menjadi faktor penting dari interaksi antar manusia dan komputer. [1]

Automatic Speech Recognition (ASR) adalah suatu sistem yang memungkinkan komputer untuk menerima masukan berupa ucapan. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami suara yang diucapkan dengan cara digitalisasi huruf dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Hasil dari identifikasi huruf yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai sebuah perintah untuk melakukan suatu pekerjaan. [2]. ASR sistem terdiri dari tiga proses yaitu fitur ekstraksi, pelatihan, dan Matching.

Proses ekstraksi ciri merupakan salah satu bagian terpenting dalam ASR, karena kemampuan ASR dalam mengenali masukan suara sangat bergantung pada ciri yang diekstraksi. Dua metode ekstraksi ciri yang sejauh ini terbukti memberikan performa terbaik bagi ASR adalah MFCC dan Wavelet. Mel Frequency Cepstrum Coefficients (MFCC) diperkenalkan oleh Davis dan Mermelstein pada tahun 1980 mengusulkan menggunakan MFCC ekstraksi fitur untuk sistem pengenalan suara [3]. Sedangkan metode Wavelet sudah banyak digunakan dalam melakukan kompresi data, klarifikasi citra dan pengolahan citra. Penggunaan metode ekstraksi ciri MFCC sampai sejauh ini masih memberikan hasil yang baik untuk sistem ASR berbasis angka. Akan tetapi metode MFCC hanya dapat bekerja pada resolusi waktu terbatas. Hal ini memungkinkan terjadinya kehilangan informasi pada suara yang diproses. Berbeda dengan MFCC, Wavelet memiliki kemampuan Multi-resolusi karena basis fungsi yang digunakan dapat bekerja pada domain waktu dan frekuensi secara bersamaan sehingga tidak terjadi kehilangan informasi.

Penelitian pengenalan suara telah banyak dilakukan, Puspitasari dkk. menggunakan MFCC dan SOM dalam pengenalan ucapan angka satu sampai sepuluh dalam bahasa Indonesia dan

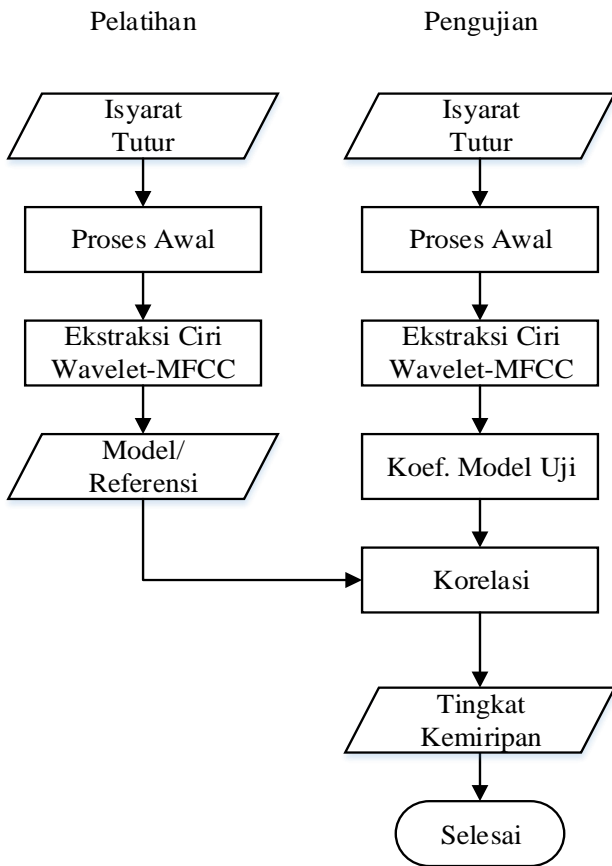
melakukan analisis faktor yang mempengaruhi akurasi dari pengenalan ucapan untuk mengembangkan sistem pengendali perangkat listrik [4]. Amalia melakukan pengenalan pola suara berupa digit 0 sampai 9 dengan menggunakan MFCC untuk proses ekstraksi dan algoritma backpropagation untuk menguji dan melatih hasil ekstraksi [5]. Sedangkan Emilia dkk. menganalisis tingkat akurasi yang dihasilkan oleh metode HMM dengan data suara dalam bahasa Indonesia yang hanya difokuskan pada pengucapan kata nol, kosong, dan satu sampai sembilan [6]. Hidayat dkk. membuat sistem pengenalan suara suku kata bahasa Indonesia dengan memanfaatkan MFCC dan Wavelet sebagai pengekstraksi ciri. Hasil menunjukkan ciri Wavelet dapat memberikan hasil pengenalan yang setara dengan ciri MFCC [7]. Hal ini dikarenakan Wavelet memiliki kelebihan dari MFCC, yaitu dapat mengolah data secara keseluruhan ke dalam domain frekuensi-waktu tanpa terjadi kehilangan data asli [7].

Berdasarkan pemaparan di atas, akan dilakukan penelitian tentang pengenalan suara digit dengan menggunakan gabungan metode ekstraksi ciri Wavelet-MFCC. Untuk pengenalan suara menggunakan metode korelasi. Perhitungan performa sistem menggunakan metode Word Recognition Rate (WRR) yang divalidasi dengan K-Fold Cross Validation.

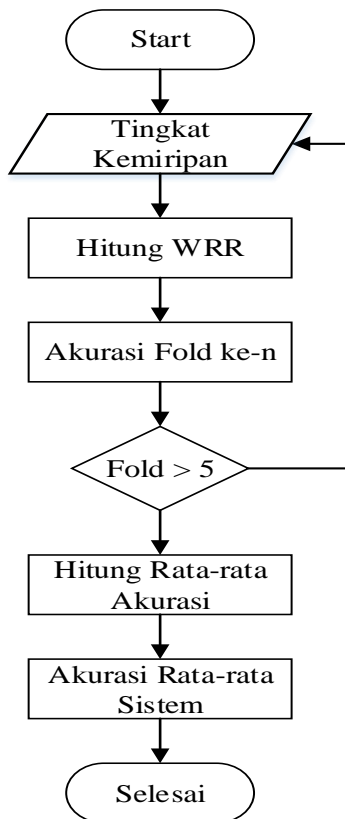
## II. METODOLOGI

Sistem pengenalan digit dengan Wavelet-MFCC ditunjukkan pada Gbr. 1. Masukan sistem berupa isyarat tutur, kemudian diterapkan pemrosesan awal. Setelah melalui tahap pemrosesan awal, suara akan diekstraksi cirinya dengan metode Wavelet-MFCC. Hasil ekstraksi kemudian dijadikan sebagai ciri model/referensi. Ciri referensi akan dipanggil kembali pada proses pengujian menggunakan metode korelasi. Hasil korelasi kemudian disimpan untuk dihitung akurasinya dan divalidasi dengan k-fold cross validation. Gambaran proses validasi ditunjukkan pada Gbr. 2.

Lebih detail tentang setiap proses yang terdapat di dalam diagram alir Gbr. 1 dan Gbr. 2 akan dijelaskan pada sub bagian selanjutnya.



Gbr. 1. Sistem Pengenal Digit dengan Wavelet-MFCC



Gbr. 2. Proses validasi sistem Pengenal Digit dengan Wavelet-MFCC

A. Dataset Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data diperoleh dari <https://github.com/Jakobovski/free-spoken-digit-dataset> (Jackson, 2019). Data berupa rekaman suara pengucapan digit nol (0) sampai sembilan (9) yang diucapkan dalam bahasa Inggris oleh empat orang pembicara laki-laki. Dua diantaranya adalah pembicara asli, sementara dua yang lain tidak. Setiap pembicara mengucapkan satu pengucapan digit sebanyak lima puluh kali. Dengan demikian jumlah total data yang digunakan adalah dua ribu sampel suara. Dalam proses perekamannya digunakan frekuensi sampling (fs) sebesar 8 kHz. Dari seluruh data yang ada 20% akan dijadikan sebagai data uji dan sisanya dijadikan sebagai data latih. Penerapan k-fold cross validation menjadikan setiap data akan menjadi dataset uji dalam setiap prosesnya.

B. Pemrosesan Awal

Terdapat dua proses pada tahap pengolahan awal penelitian ini yaitu, normalisasi dan penapisan dengan preemphasis.

1) Normalisasi

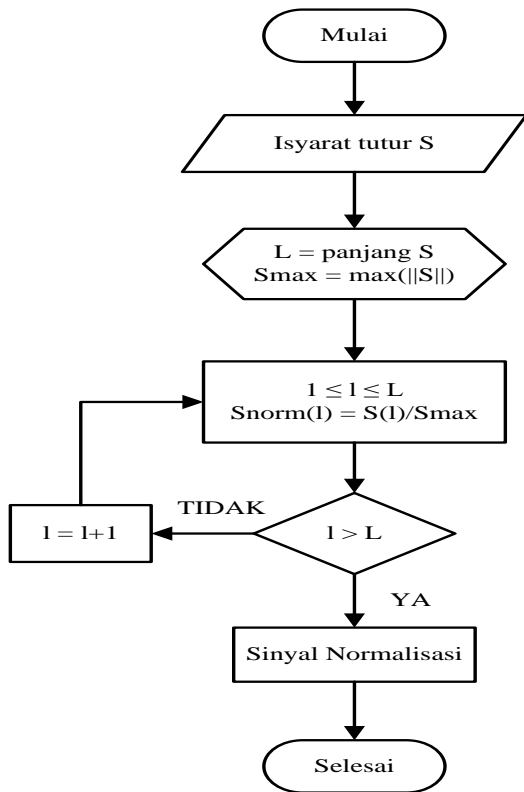
Normalisasi bertujuan untuk menyeragamkan nilai amplitudo isyarat suara sehingga amplitudo maksimal masing-masing sampel isyarat suara adalah -1 dan 1. Untuk memperoleh isyarat ternormalisasi yaitu dengan membagi tiap-tiap sampel suara dengan nilai maksimum harga mutlak sampel suara tersebut, berdasarkan persamaan (1). Pada Gbr. 3. ditunjukkan diagram alir proses normalisasi.

$$S_{no}(n) = \frac{S(n)}{\max(abs(S(n)))} \tag{1}$$

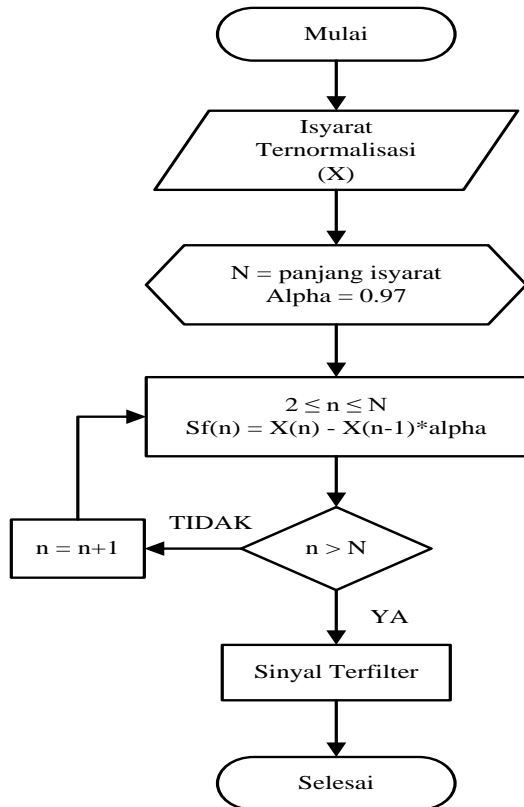
2) Preemphasis

Untuk membuat tapis preemphasis merujuk pada persamaan (2). Tapis ini diaplikasikan untuk memperbaiki isyarat dari gangguan derau sehingga akan meningkatkan kualitas suara dan ciri yang diekstraksi. Nilai alfa  $\alpha$  yang digunakan pada proses ini adalah sebesar 0.97, berdasarkan percobaan awal [8]. Prosesnya yaitu mengurangi nilai sampel ke-n dengan dirinya sendiri yang telah dikalikan dengan  $\alpha$ . Pada Gbr. 4. ditunjukkan diagram alir sistem preemphasis.

$$y[n] = x[n] - ax[n - 1] \quad (2)$$



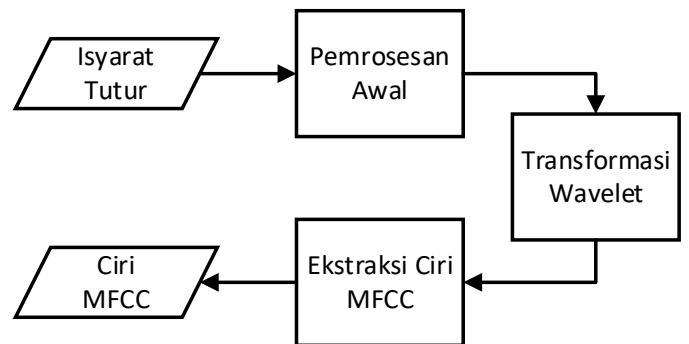
Gbr. 3. Diagram Alir Proses Normalisasi



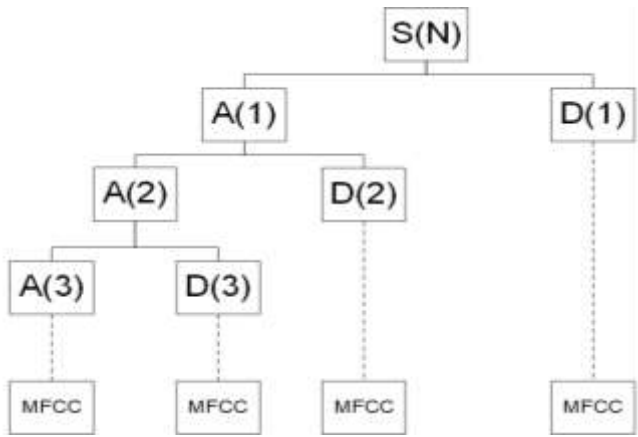
Gbr. 4. Diagram alir proses Tapis Preemphasis

C. Ekstraksi Ciri Wavelet-MFCC

Gambaran umum proses ekstraksi ciri Wavelet-MFCC ditunjukkan pada Gbr. 5. Isyarat tutur sebagai masukan pada tahap pemrosesan awal akan dinormalisasi dan dihilangkan deraunya dengan proses preemphasis. Selanjutnya, isyarat tutur yang telah melalui pemrosesan awal akan diimplementasikan transformasi wavelet kepadanya. Pada proses transformasi ini terdapat dua jenis metode dekomposisi yang diterapkan yaitu dekomposisi wavelet dyadic dan dekomposisi wavelet packet. Pada dekomposisi wavelet dyadic, hanya kanal Aproksimasi (cA) yang akan didekomposisi untuk level selanjutnya, sedangkan pada proses dekomposisi wavelet packet baik kanal Aproksimasi (cA) maupun kanal Detail (cD) akan didekomposisi pada level selanjutnya. Selanjutnya hasil dekomposisi pada level terakhir yang akan diubah menjadi koefisien ciri MFCC. Gambaran proses ekstraksi ciri wavelet secara detail baik untuk dekomposisi wavelet dyadic dan wavelet packet ditunjukkan pada Gbr. 6 dan Gbr. 7. Pada kedua gambar dekomposisi dilakukan sampai level 3.

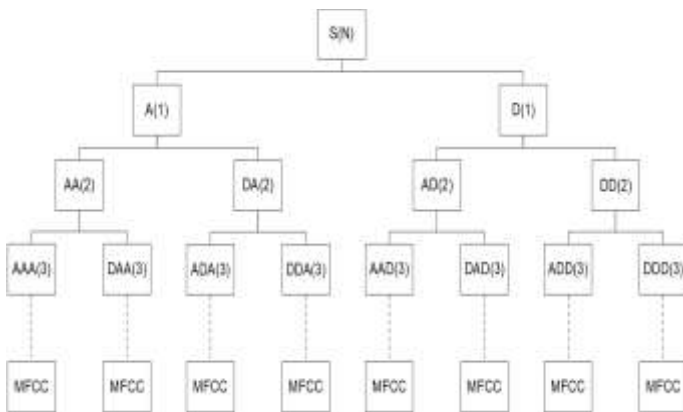


Gbr. 5. Proses ekstraksi ciri Wavelet-MFCC



Gbr. 6. Skema Pembentukan Ciri MFCC Wavelet Dyadic pada dekomposisi level 3

Pada penelitian ini, untuk proses ekstraksi ciri Wavelet-MFCC, digunakan wavelet Daubechies db3, db5, dan db7. Pemilihan wavelet Daubechies seperti yang telah dilakukan dalam beberapa penelitian yang telah lalu seperti, [9], [10] yang telah memanfaatkan wavelet family Daubechies pada sinyal suara fonem dan suku kata. Sedangkan untuk level dekomposisi yang diimplementasikan adalah level 6. Sehingga pada proses pembentukan ciri MFCC untuk wavelet dyadic terdapat 7 kanal yang digunakan, yaitu cA6, cD6, cD5, cD4, cD3, cD2, dan cD1. Sementara itu untuk hasil dekomposisi wavelet packet jumlah kanal yang diproses lebih banyak, yaitu sebanyak  $2^6 = 64$  kanal pada pembentukan ciri MFCC.



Gbr. 7. Skema Pembentukan Ciri MFCC Wavelet Packet pada dekomposisi level 3

Pada proses MFCC diterapkan frame dengan panjang 25 milidetik dengan overlapping sebesar 10 milidetik. Fungsi window yang digunakan adalah window Hamming. Ukuran data pada proses Fast Fourier Transform (FFT) adalah 512. Sementara jumlah cepstrum yang dibentuk adalah 13 yang

diambil dari proses wrapping dengan filterbank yang memiliki filter segitiga sebanyak 26. Sehingga, masing-masing digit setelah proses ekstraksi ciri MFCC akan memiliki panjang runtun ciri yang berbeda. Untuk ciri wavelet-MFCC dengan wavelet dyadic memiliki panjang runtun 91 ciri, sedangkan untuk wavelet packet panjang runtunnya 832 ciri. Jumlah ini diperoleh dari pengambilan hanya 13 dari total 26 koefisien MFCC yang dijadikan sebagai ciri dari setiap kanal hasil dekomposisi.

D. Pengenalan

Metode korelasi diimplementasikan pada proses pengenalan dalam penelitian ini. Metode korelasi adalah satu penelitian untuk mengetahui hubungan dan tingkat hubungan antara dua variabel atau lebih tanpa ada upaya untuk mempengaruhi variabel tersebut sehingga tidak terdapat manipulasi variabel. Dalam penerapannya akan dicari nilai koefisien korelasi. Koefisien korelasi adalah ukuran statistik yang menghitung kekuatan hubungan antara gerakan relatif dua variabel. Nilai berkisar antara -1 dan 1. Angka yang dihitung lebih besar dari 1 atau kurang dari -1 berarti bahwa ada kesalahan dalam pengukuran korelasi. Korelasi -1 menunjukkan korelasi negatif sempurna, sedangkan korelasi 1 menunjukkan korelasi positif sempurna. Korelasi 0 menunjukkan tidak ada hubungan antara pergerakan kedua variabel. Koefisien korelasi diukur dengan persamaan (3):

$$\rho_{xy} = \frac{Cov(r_x, r_y)}{\sigma_x \sigma_y} \tag{3}$$

Didalam penelitian ini nilai koefisien yang digunakan berkisar antar 0 dan 1[11].

E. Word Error Rate (WER)

Word Error Rate adalah metrik umum dari kinerja pengenalan suara atau sistem terjemahan mesin. Kesulitan umum mengukur kinerja terletak pada kenyataan bahwa urutan kata yang dikenali dapat memiliki panjang yang berbeda dari urutan kata referensi (seharusnya yang benar). WER diturunkan dari jarak Levenshtein, bekerja pada level kata dan bukannya level fonem. Masalah ini diselesaikan dengan terlebih dahulu menyelaraskan urutan kata yang dikenali dengan urutan kata referensi

(diucapkan) menggunakan penyelarasan string dinamis[12], [13]. Tingkat kesalahan kata kemudian dapat dihitung sebagai:

$$W E R = \frac{S + D + I}{N} \times 100\% \quad (4)$$

Di mana *S* adalah jumlah kata yang tersubstitusi, *D* adalah jumlah kata yang terhapus, *I* adalah jumlah sisipan kata, dan *N* adalah jumlah kata dalam referensi. Saat melaporkan kinerja sistem pengenalan ucapan, dapat digunakan tingkat pengenalan kata Word Recognition Rate (WRR). WRR diturunkan dari persamaan WER yang dinyatakan sebagai:

$$W R R = 1 - W E R = \frac{N - S - D - I}{N} = \frac{C - I}{N} \quad (5)$$

dengan  $C = N - (S + D)$ , jumlah kata-kata yang dikenali dengan benar.

F. Validasi K-Fold Cross Validation

.Untuk menjaga agar variasi hasil pengenalan tetap rendah, terkait dengan pemilihan data latih dan data uji, dilakukan K fold cross validation. Pada teknik ini, data dibagi sejumlah K dataset. Dataset ke-1 digunakan sebagai data uji sedangkan dataset ke-2, 3, ..., K digunakan sebagai data latih. Kemudian, hasil akurasi dicatat. Selanjutnya proses diulangi dengan dataset ke-2 sebagai data uji dan dataset ke-1, 3, ..., K digunakan sebagai data latih. Proses diulangi sampai sejumlah K kali dan akurasi total adalah jumlah akurasi dibagi K [14], [15].

Penerapan *k-fold cross validation* pada *dataset* dalam penelitian ini menggunakan jumlah *fold* 5. Sehingga pada setiap *fold* digunakan 20% data sebagai *dataset* uji dan sisanya sebagai *dataset* latih. Untuk *fold* pertama 20 persen data pertamanya adalah data ke-1 sampai data ke-400 yang digunakan sebagai data uji untuk mendapatkan akurasi sistem pada *fold* pertama tersebut. Begitu seterusnya sehingga pada *fold* ke-5 digunakan data ke-1601 sampai data ke-2000 untuk mendapatkan akurasi sistem pada *fold* kelima. Kemudian hasil seluruh akurasi dari fold ke-1 hingga fold ke-5 dihitung rata-ratanya untuk mendapatkan akurasi sistem yang valid. Ilustrasi pembagian data pada setiap fold ditunjukkan pada tabel dibawah ini.

TABEL I  
PENERAPAN K-FOLD CROSS VALIATION DENGAN JUMLAH FOLD 5, DATASET LATIH 20% DAN DATASET UJI 80%.

FOLD Ke	JUMLAH DATASET (2000 data)				
	20%	40%	60%	80%	100%
1	1-400				
2		401-800			
3			801-1200		
4				1201-1600	
5					1601-2000

Keterangan:

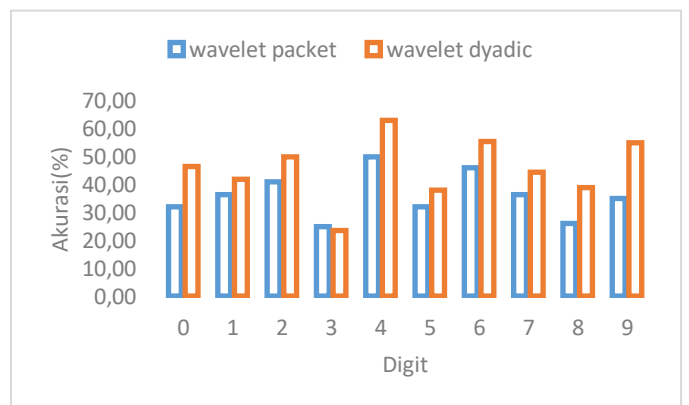
- : dataset uji
- : dataset latih

Analisis terhadap penerapan wavelet Daubechies db3, db5, dan db7 dan penggunaan transformasi jenis dyadic dan packet akan dilakukan untuk mengetahui akurasi terbaik sistem dan metode dekomposisi wavelet yang tepat untuk penelitian pengenalan suara digit ini.

III. HASIL DAN PEMBAHASAN

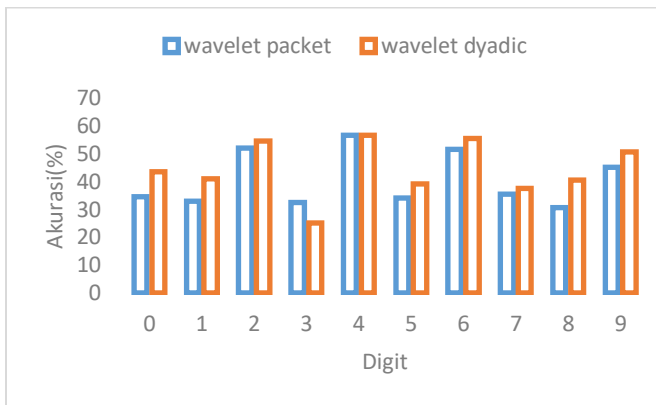
A. Akurasi Untuk db3

Hasil akurasi untuk penerapan wavelet db3 pada pengenalan suara digit 0 sampai 9 dengan menggunakan metode dekomposisi wavelet packet dan wavelet dyadic dapat dilihat pada Gbr. 8. Akurasi pengenalan suara digit 0 sampai 9 lebih baik menggunakan wavelet dyadic daripada wavelet packet, dimana selisih akurasi keduanya dapat mencapai lebih dari 5%. Pada akurasi menggunakan db3, hasil pembacaan suara tertinggi berada pada data suara digit 4 dengan menggunakan wavelet dyadic dengan akurasi 63%, sedangkan data suara digit 4 dengan menggunakan wavelet packet dengan akurasi 50%.



Gbr. 8. Akurasi Daubechies db3

### B. Akurasi Untuk db5

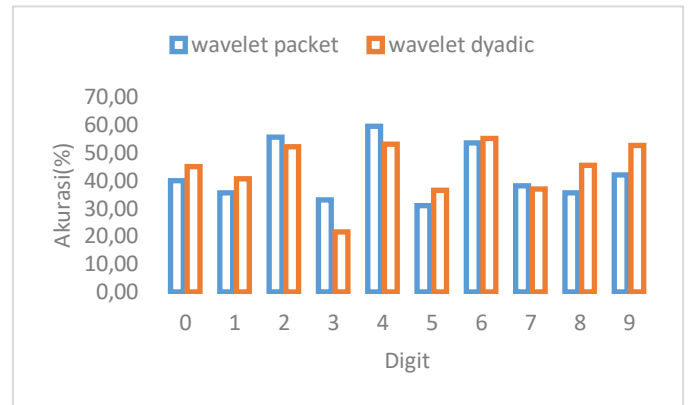


Gbr. 9. Akurasi Daubechies db5

Hasil akurasi untuk penerapan wavelet db5 pada pengenalan suara digit 0 sampai 9 dengan menggunakan dekomposisi wavelet packet dan wavelet dyadic dapat dilihat pada Gbr. 9. Pada akurasi menggunakan db5, hasil pengenalan suara tertinggi berada pada data suara digit 4 menggunakan wavelet dyadic dan wavelet packet dengan akurasi yang sama, yaitu 56,5%. Meskipun demikian, secara umum untuk akurasi pengenalan suara digit 0 sampai 9 masih lebih baik menggunakan wavelet dyadic daripada wavelet packet.

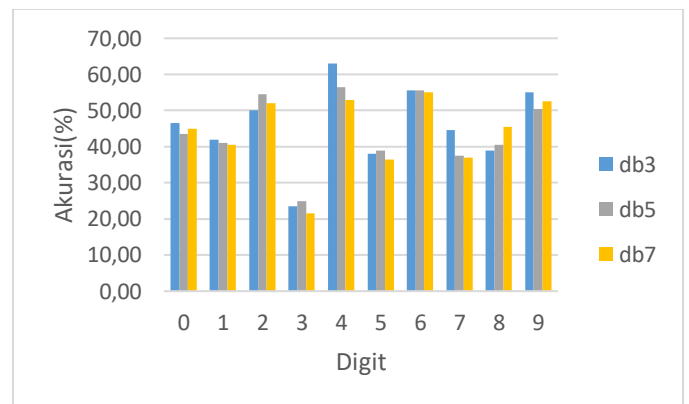
### C. Akurasi Untuk db7

Hasil akurasi untuk penerapan wavelet db7 pada pengenalan suara digit 0 sampai 9 dengan menggunakan dekomposisi wavelet packet dan wavelet dyadic dapat dilihat pada Gbr. 10. Pada akurasi dengan menggunakan db7, hasil pengenalan suara tertinggi berada pada data suara digit 4 menggunakan wavelet packet dengan akurasi 59,5% sementara dengan wavelet dyadic akurasinya 53%. Meskipun demikian pengenalan suara digit 0 sampai 9 dengan menggunakan wavelet dyadic masih lebih dominan dibandingkan menggunakan wavelet packet.



Gbr. 10. Akurasi Daubechies db7

### D. Akurasi Menggunakan Wavelet Dyadic



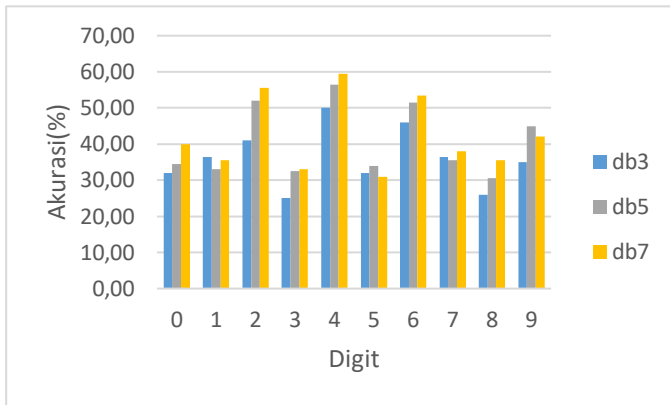
Gbr. 11. Akurasi Rata-rata Implementasi Wavelet Dyadic

Hasil akurasi rata-rata penerapan wavelet dyadic dapat dilihat pada Gbr. 11. Akurasi tertinggi didominasi oleh db3, kemudian db5 dan terakhir db7. Hasil pengenalan suara menggunakan wavelet dyadic mengalami penurunan akurasi yang dipengaruhi oleh peningkatan orde. Sebagaimana diketahui proses dekomposisi wavelet dyadic hanya diterapkan pada koefisien aproksimasi (A). Sehingga penurunan akurasi dapat dipengaruhi oleh proses ini.

### E. Akurasi Menggunakan Wavelet Packet

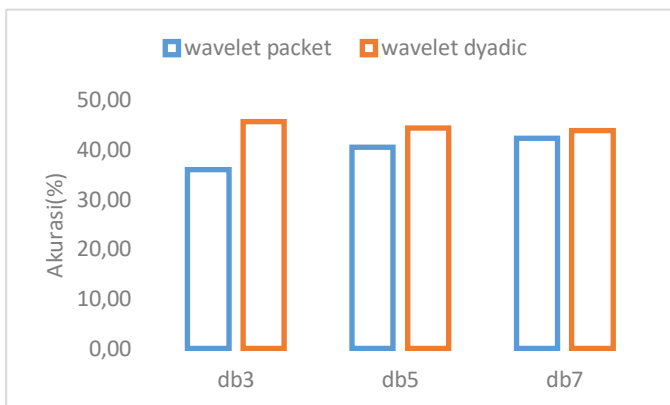
Hasil akurasi rata-rata penerapan wavelet packet dapat dilihat pada gambar 4.25. Akurasi tertinggi didominasi oleh db7, kemudian db5 dan terakhir db 3. Hasil pengenalan suara menggunakan wavelet packet mengalami peningkatan akurasi yang dipengaruhi oleh penurunan orde. Sebagaimana diketahui proses dekomposisi wavelet packet diterapkan pada koefisien aproksimasi (A) dan dekomposisi (D). Sehingga peningkatan akurasi dapat dipengaruhi oleh proses ini.





Gbr. 12. Akurasi Rata-rata Impelementasi Wavelet Packet

F. Akurasi Perbandingan WP dan WD



Gbr. 13. Perbandingan Akurasi Rata-rata Wavelet Packet dan Wavelet Dyadic

Hasil akurasi perbandingan wavelet packet dan wavelet dyadic dapat dilihat pada Gbr. 13. Pada perbandingan akurasi menggunakan wavelet packet dan wavelet dyadic, wavelet packet mengalami kenaikan sedangkan wavelet dyadic mengalami penurunan. Meskipun demikian nilai akurasi terendah dari wavelet dyadic masih lebih tinggi daripada akurasi tertinggi dari wavelet packet. Oleh karena itu untuk pengenalan suara digit 0 sampai 9 lebih baik menggunakan wavelet dyadic daripada menggunakan wavelet packet.

IV. PENUTUP

Penerapan wavelet Daubechies dalam system pengenalan suara digit 0-9 telah dilakukan. Wavelet daubechies yang digunakan adalah db3, db5, dan db7. Adapun metode transformasi wavelet yang diterapkan adalah transformasi wavelet dyadic dan

wavelet packet dengan level dekomposisi ke 6. Pengenalannya menggunakan korelasi yang divalidasi dengan menghitung WER dan K-fold Validation.

Dari seluruh hasil yang telah ditampilkan pada subbab 4.3. sampai 4.8. didapatkan hasil bahwa untuk pengenalan digit yang terbaik setelah diterapkannya metode ini adalah digit yang dapat dikenali paling baik adalah digit 4 dengan akurasi tertingginya 63 % pada wavelet dyadic db3. Sedangkan untuk penggunaan metode transformasi antara wavelet dyadic dan wavelet packet, diketahui bahwa dalam penelitian ini penggunaan wavelet dyadic masih lebih baik dari penerapan wavelet packet. Namun jika melihat dari tren grafik yang ditunjukkan pada gambar 4.26. dapat diprediksi akan adanya peningkatan penerapan transformasi wavelet packet untuk peningkatan orde waveletnya. Namun dibutuhkan penelitian lebih jauh untuk dapat membuktikannya.

Secara umum, akurasi yang diperoleh pada penelitian ini dapat dipengaruhi oleh banyak faktor, di antaranya dataset yang digunakan, algoritma pengenalan dan metode validasi yang diterapkan. Untuk dataset yang digunakan, karena dataset ini merupakan dataset sekunder, maka variable luar yang mempengaruhinya sangatlah besar. Sehingga untuk meminimalisir pengaruh variable tersebut diterapkan metode k-fold validation. Hasil menunjukkan bahwa penggunaan dataset ini dapat memberikan akurasi maksimal sekitar 63 %, khusus pada penelitian ini.

Pengaruh kedua adalah algoritma pengenalan, koefisien korelasi, yang diterapkan. Algoritma ini sebenarnya sangat baik untuk diterapkan jika menggunakan ciri statistik dan dengan jumlah data yang tidak terlalu besar. Akurasi yang rendah setelah menerapkan algoritma koefisien korelasi dapat disebabkan karena fitur ciri yang diperoleh dari penerapan wavelet-MFCC dapat dikatakan sangat besar. Seperti pada ekstraksi wavelet packet, setiap model memiliki fitur sebanyak 832 ciri. Sehingga jika diterapkan korelasi terhadapnya akan mempengaruhi variabilitasnya. Padahal ciri yang baik haruslah memiliki variabilitas yang tinggi antar satu ciri dengan ciri yang lainnya.

Penelitian ke depannya dapat difokuskan pada peningkatan akurasi sistem misalkan dengan mengaplikasikan algoritma Jaringan Syaraf Tiruan (JST). Selain itu, masih dimungkinkan untuk



menerapkan jenis wavelet family yang lain dengan orde yang lebih tinggi untuk mengeksplorasi jenis wavelet terbaik pada pengembangan sistem pengenalan suara digit.

#### REFERENSI

- [1] R. L. K. Venkateswarlu, R. Raviteja, and R. Rajeev, "The Performance Evaluation of Speech Recognition by Comparative Approach," in *Advances in Data Mining Knowledge Discovery and Applications estimating*, vol. i, no. tourism, 2016, p. 13.
- [2] J. M. T. S., D. Puspitaningrum, and B. Susilo, "Penerapan Speech Recognition Pada Permainan Teka-Teki Silang Menggunakan Metode Hidden Markov Model (HMM) Berbasis Desktop," *J. Rekrusif*, vol. 4, no. 1, pp. 119–129, 2016.
- [3] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoust.*, vol. 28, no. 4, pp. 357–366, 1980, doi: 10.1109/TASSP.1980.1163420.
- [4] M. Puspitasari, J. Supardi, and Y. Sazaki, "PENGENALAN SUARA MENGGUNAKAN MEL FREQUENCY CEPSTRAL COEFFICIENTS DAN SELF ORGANIZING MAPS," *ePrints Sriwij. Univ.*, 2014, [Online]. Available: [http://eprints.unsri.ac.id/5493/1/PENGENALAN\\_SUARA\\_MENGGUNAKAN\\_MEL\\_FREQUENY\\_CEPSTRAL\\_COEFFICIENTS\\_DAN\\_SELF\\_ORGANIZING\\_MAPS.pdf](http://eprints.unsri.ac.id/5493/1/PENGENALAN_SUARA_MENGGUNAKAN_MEL_FREQUENY_CEPSTRAL_COEFFICIENTS_DAN_SELF_ORGANIZING_MAPS.pdf).
- [5] S. Amalia, "Pengenalan Digit 0 Sampai 9 Menggunakan Ekstraksi Ciri MFCC dan Jaringan Syaraf Tiruan Backpropagation," *J. Tek. Elektro ITP*, vol. 6, no. 1, pp. 1–8, 2017, doi: 10.21063/jte.2017.3133601.
- [6] N. R. Emillia, Suyanto, and W. Maharani, "Penggunaan Hidden Markov Model Dan Genetic Algorithm Untuk Pemodelan Automatic Speech Recognition Pada Pengenalan Ucapan Bahasa Indonesia ( Speech To Text ) Usage of Hidden Markov Models and Genetic Algorithm for Automatic Speech," Universitas Telkom, 2011.
- [7] S. Hidayat, R. Hidayat, and T. B. Adji, "Sistem Pengenal Tutar Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM," in *CITEE*, 2015, no. September, pp. 246–251.
- [8] Abriyono and A. Harjoko, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *IJCCS*, vol. 6, no. 2, pp. 23–34, 2012.
- [9] O. Farooq and S. Datta, "Phoneme recognition using wavelet based features," *Inf. Sci. (Ny)*, vol. 150, no. 1–2, pp. 5–15, 2003, doi: 10.1016/S0020-0255(02)00366-3.
- [10] S. Hidayat, R. Hidayat, and T. B. Adji, "Speech Recognition of Kv-Patterned Indonesian Syllable Using Mfcc, Wavelet and Hmm," *Kursor*, vol. 8, no. 2, p. 67, 2015, doi: 10.28961/kursor.v8i2.63.
- [11] S. Hidayat, Abdurahim, and M. Tajuddin, "Evaluation and design of wavelet packet cepstral coefficient (WPCC) for a noisy Indonesian vowels signal," *J. Phys. Conf. Ser.*, vol. 1211, no. 1, 2019, doi: 10.1088/1742-6596/1211/1/012023.
- [12] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A Review on Speech Recognition Technique," *Int. J. Comput. Appl.*, vol. 10, no. 3, pp. 16–24, 2010, doi: 10.5120/1462-1976.
- [13] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall, 2001.
- [14] A. Rizal, "Perbandingan Skema Dekomposisi Paket Wavelet untuk Pengenalan Sinyal EKG," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 4, no. 2, 2015, doi: 10.22146/jnteti.v4i2.145.
- [15] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.