

Ekstraksi Ciri dan Klasifikasi Isyarat Suara Tutar Menggunakan Metode Mel Frequency Cepstral Coefficients

Domy Kristomo^{a1}, Maria Mediadrix Sebatubun^a

^aTeknologi Informasi, Universitas Teknologi Digital Indonesia
Jalan Raya Janti 143 Karang Jambe Yogyakarta 55198, Indonesia
¹domy@utdi.ac.id

Abstract

Teknologi *speech recognition* telah mengalami kemajuan signifikan dalam beberapa dekade terakhir, didorong oleh perkembangan dalam bidang *Artificial Intelligence (AI)*, *Machine Learning (ML)*, dan *Natural Language Processing (NLP)*. Teknologi ini memungkinkan sistem komputer untuk mengenali, memahami, dan memproses ucapan manusia secara langsung, sehingga menciptakan interaksi manusia-mesin yang lebih intuitif dan efektif. Berbagai metode telah digunakan dalam pemrosesan dan ekstraksi ciri isyarat tutur salah satu diantaranya adalah *Mel-frequency cepstral coefficients (MFCC)*. Penelitian ini menggunakan metode MFCC untuk ekstraksi ciri dan *Multilayer Perceptron (MLP)* untuk klasifikasi isyarat suara tutur. MFCC, yang didasarkan pada skala Mel, dirancang untuk merepresentasikan karakteristik frekuensi isyarat suara sesuai dengan persepsi pendengaran manusia. Proses ekstraksi melibatkan tahapan *pre-emphasis*, *framing*, *windowing*, transformasi *Fourier*, filter bank Mel, dan transformasi kosinus terdiskret (*DCT*) untuk menghasilkan fitur suara yang kompak dan informatif. Fitur ini kemudian digunakan sebagai input dalam model klasifikasi berbasis pembelajaran mesin. Hasil pengujian menunjukkan bahwa metode MFCC mampu menangkap pola suara dengan akurasi yang tinggi, bahkan dalam kondisi dengan gangguan *noise*. Studi ini menunjukkan bahwa kombinasi MFCC dengan algoritma klasifikasi MLP dapat meningkatkan performa sistem pengenalan suara tutur secara signifikan, ditunjukkan oleh hasil akurasi klasifikasi sebesar 94,33%.

Keywords: *Mel Frequency Cepstral Coefficients, Ekstraksi Ciri, Klasifikasi, Speech, Word*

1. Pendahuluan

Teknologi *speech recognition* atau pengenalan suara tutur telah berkembang pesat dalam beberapa dekade terakhir, didukung oleh kemajuan di bidang kecerdasan buatan (*AI*), pembelajaran mesin (*ML*), dan pemrosesan bahasa alami (*NLP*). Teknologi ini memungkinkan komputer untuk mengenali, memahami, dan menginterpretasikan ucapan manusia secara *real-time*, membuka jalan bagi interaksi manusia-mesin yang lebih alami dan efisien. Dari penggunaannya dalam perangkat asisten *virtual* seperti *Siri*, *Alexa*, dan *Google Assistant*, hingga penerapannya dalam layanan kesehatan, pendidikan, dan otomotif, *speech recognition* telah menjadi bagian integral dari kehidupan modern. Dengan terus meningkatnya akurasi dan kemampuan multibahasa, teknologi ini menawarkan potensi besar untuk mengatasi hambatan komunikasi dan mendukung inklusivitas di berbagai sektor. Berbagai metode telah digunakan dalam pemrosesan dan ekstraksi ciri isyarat tutur diantaranya *Mel-frequency cepstral coefficients (MFCC)*, *Wavelet*, *Linear Predictive Coefficients*, dan sebagainya.

Mel-frequency cepstral coefficients (MFCC) adalah salah satu metode yang paling populer dan efektif untuk ekstraksi ciri dalam analisis isyarat suara, termasuk isyarat tutur. Metode ini dirancang untuk menangkap karakteristik frekuensi yang relevan dengan persepsi pendengaran manusia, menggunakan konsep skala mel untuk merepresentasikan frekuensi secara lebih sesuai dengan cara telinga manusia memproses suara. Dalam prosesnya, MFCC melibatkan langkah-langkah seperti transformasi *Fourier*, penggunaan filter bank mel, dan penerapan transformasi kosinus terdiskret (*DCT*)

untuk menghasilkan representasi numerik yang kompak dari isyarat suara. Karena keandalannya dalam menggambarkan ciri akustik yang penting, MFCC telah diterapkan dalam berbagai aplikasi seperti pengenalan suara, verifikasi pembicara, dan analisis emosi dari isyarat suara.

Penelitian sebelumnya terkait penerapan MFCC dalam pengolahan isyarat suara, telah banyak dilakukan [1][2][3][4][5][6][7][8][9][10]. Paleva dan Prasetio [1] menggunakan *Short Time Fourier Transform* (STFT) pada ciri MFCC untuk mendeteksi tingkat stres melalui suara pengguna menggunakan perangkat berbasis Android. Penelitian ini menunjukkan peningkatan akurasi sistem dari 66% menggunakan MFCC standar menjadi 73% dengan penerapan STFT pada MFCC. Ajinurseto dkk. [2] menerapkan MFCC pada sistem pengenalan suara berbasis desktop sebagai bagian dari teknologi biometrik suara. Sistem tersebut diuji dalam dua kondisi, yaitu ideal (tanpa *noise*) dan tidak ideal (dengan *noise*). Hasil pengujian menunjukkan akurasi sistem mencapai 90% dalam kondisi ideal dan 76,67% dalam kondisi tidak ideal. Santoso dkk. [3] menggunakan MFCC untuk ekstraksi ciri sembilan kata perintah dengan tiga variasi intonasi, melibatkan 100 orang untuk menghasilkan dataset suara. Proses pengolahan melibatkan tahap seperti *hamming window*, transformasi Fourier cepat (FFT), dan transformasi Fourier terbalik (IFFT) untuk menghasilkan spektrum, spektrogram, log-spektrogram, dan koefisien MFCC. Studi ini menunjukkan bahwa penggunaan rentang frekuensi 50-1635 Hz menghasilkan akurasi terbaik dalam menghilangkan *noise* dan mengekstraksi fitur suara. Buana dkk. [4] menggunakan MFCC dan Modified K-Nearest Neighbor untuk prediksi suara paru dengan hasil akurasi tertinggi pada Fold 4 yaitu 0.98 atau 98% dengan precision 1, recall 0.14, f1-score 0.25 untuk normal dan precision 0.97, recall 1, f1-score 0.98 untuk tidak normal sebesar 97,83 % dengan menggunakan metode uji dan data latih yang telah disediakan. Sasilo dkk. [5] menggunakan MFCC dan *Gaussian Mixture Model* (GMM) untuk pengenalan suara. Berdasarkan hasil pengujian, metode MFCC - GMM pada kondisi ideal memiliki tingkat akurasi sebesar 82.22% sedangkan pada kondisi tidak ideal mendapatkan akurasi sebesar 66.67%. Tridarma dan Endah [6] menggunakan MFCC dan *Recurrent Neural Network* untuk pengenalan ucapan Bahasa Indonesia. Hidayat dan Winursito [7] menggunakan MFCC untuk ekstraksi fitur dan metode *K-nearest neighbour* (KNN) untuk klasifikasi. Tiga versi dari metode yang diusulkan dirancang. Hasilnya menunjukkan bahwa versi tiga meningkatkan akurasi sebesar 4% dibandingkan sistem pengenalan konvensional. Versi ketiga memiliki waktu komputasi paling cepat dibandingkan metode lainnya. Penambahan algoritma deteksi suku kata pada versi tiga meningkatkan waktu komputasi hanya 0,151 detik dibandingkan metode MFCC konvensional. Anggoro dkk. [8] melakukan identifikasi bahasa *Cleft lip and palate* (CLP) dan pola bunyi standar menggunakan ekstraksi fitur MFCC dengan klasifikasi *K-Nearest Neighbor* (KNN) dan validasi silang K-Fold. Dengan menjadikan kata yang mempunyai huruf /p/ sebagai acuannya disebut bilabial. Kata-kata yang digunakan antara lain Paku, Kapak, dan Atap. Keakuratan hasil pengenalan mencapai lebih dari 69%, dengan akurasi minimal 41%. Prabakaran dan Sriuppili [9] menyelidiki skema ekstraksi fitur berdasarkan MFCC. Permana dkk. [10] menggunakan MFCC dan DTW (*Dynamic Time Warping*) untuk pengenalan jenis suara pria dan Wanita. Hasil Penelitian yang didapat yaitu untuk tingkat akurasi pada wanita dengan jenis suara alto didapatkan presentase 80%, untuk tingkat akurasi jenis suara mezzosopran didapatkan 90%, untuk tingkat akurasi jenis suara sopran didapatkan 80%. Kemudian pada jenis suara pria, untuk jenis suara bass didapatkan tingkat akurasi sebesar 80%, untuk jenis suara bariton didapatkan tingkat akurasi 70%, dan untuk jenis suara tenor didapatkan tingkat akurasi sebesar 60%.

Pada penelitian ini digunakan metode ekstraksi ciri MFCC dan pengklasifikasi *Multilayer-perceptron* untuk mengklasifikasi data isyarat suara kata yang mengandung *stop consonant* dalam kata tersebut dengan sedikit *noise*. *Stop consonant* terdiri atas *voiceless /p/, /t/, /k/*, dan *voiced /b/, /d/, /g/*. Data kata tersebut yaitu kakak, tutup, bibit, papan, duduk, gigit.

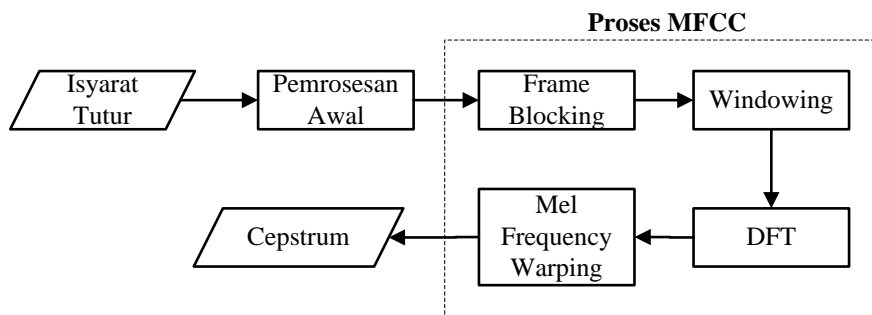
2. Metode Penelitian

Metodologi yang diusulkan dalam penelitian ini terdiri dari lima langkah. yaitu Data set, Ekstraksi ciri MFCC, Klasifikasi, validasi silang K-Fold dan akurasi.

2.1. Mel-Frequency Cepstral Coefficients (MFCC)

MFCC merupakan sebuah representasi spektrum daya waktu singkat dari isyarat suara, berdasar pada transformasi kosinus linier dari sebuah log spektrum daya pada frekuensi tak linier dalam skala Mel. Pada MFCC, band frekuensi terbagi sama besar pada skala Mel, yang mendekati persepsi sistem pendengaran manusia. Tahapan proses ekstraksi MFCC secara ringkas ditunjukkan pada Gambar 1.

Terdapat empat proses utama yaitu frame blocking, penjendelaan, transformasi fourier dan *mel frequency warping*.



Gambar 1. Proses MFCC

- *Frame Blocking*, isyarat suara merupakan isyarat non-stasioner, artinya sifat-sifat statistiknya selalu berubah terhadap waktu. Sehingga tidaklah memungkinkan mengekstraksi ciri spektral dari suara tutur sekaligus. Oleh karena itu, ciri spektral isyarat suara diekstrak melalui sebuah jendela isyarat suara yang mencirikan bagian suara tertentu sehingga dengannya dapat dibuat suatu asumsi bahwa isyarat suara tersebut stasioner. Fungsi jendela yang digunakan adalah yang bernilai maksimum 1 untuk daerah di dalam jendela dan nol untuk daerah yang lain. Jendela bergerak sepanjang isyarat suara dan mengekstraksi bentuk isyarat yang berada di dalamnya. Proses penjendelaan dicirikan oleh tiga parameter, yaitu lebar jendela, offset antar jendela dan bentuk jendela. Hasil penjendelaan disebut sebagai frame, dengan panjang (frame size) dan besar pergeseran (frame shift) tertentu dalam satuan milidetik. Rentang pergeseran frame berkisar antara 1/3 – 1/2 kali panjang frame.
- *Windowing*, Setiap frame diterapkan fungsi *windowing*, seperti *Hamming Window*, untuk mereduksi efek diskontinuitas di ujung frame. Windowing dilakukan menggunakan rumus:

$$H(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (1)$$

di mana N adalah jumlah sampel dalam frame.

- *DFT*, Frame dalam domain waktu ditransformasikan ke domain frekuensi menggunakan FFT. Hasilnya adalah spektrum frekuensi yang merepresentasikan intensitas energi dari setiap komponen frekuensi dalam frame.
- *Mel-Frequency Wrapping*, Spektrum frekuensi diubah ke skala Mel menggunakan filter bank Mel. Filter ini menggunakan frekuensi rendah yang linier dan frekuensi tinggi yang logaritmik, dengan persamaan:

$$\text{mel}(f) = 1125 \ln\left(1 + \frac{f_{\text{Hz}}}{700}\right) \quad (2)$$

Filter bank menghasilkan energi pada berbagai rentang frekuensi yang relevan dengan pendengaran manusia.

- *Cepstrum*, Nama Cepstrum diturunkan dari kata spectrum yaitu dengan membalik empat kata pertama dari spectrum. Cepstrum adalah transformasi fourier dari logaritma dengan unwrapped fase dari transformasi fourier. Cepstrum nyata menggunakan informasi magnitude dari spectrum sedangkan cepstrum kompleks mencakup magnitude dan fase dari spectrum awal. Untuk menentukan sifat-sifat cepstrum kompleks adalah dengan memperhitungkan perbandingan transformasi z yang bentuk umumnya adalah sebagai berikut:

$$X(z) = \frac{Az^r \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_0} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1}) \prod_{k=1}^{N_0} (1 - d_k z)} \quad (3)$$

2.2. Multilayer Perceptron

Multilayer Perceptron (MLP) adalah jaringan saraf tiruan yang digunakan dalam penelitian ini untuk tugas klasifikasi data. MLP terdiri dari tiga komponen utama: lapisan masukan, beberapa lapisan tersembunyi, dan lapisan keluaran. Data masukan berupa vektor fitur yang telah diekstraksi menggunakan metode *Mel-Frequency Cepstral Coefficients* (MFCC), kemudian dinormalisasi agar berada dalam rentang tertentu untuk meningkatkan efisiensi pelatihan. Arsitektur jaringan mencakup

sejumlah neuron di setiap lapisan tersembunyi, dengan fungsi aktivasi seperti ReLU pada lapisan tersembunyi dan softmax pada lapisan keluaran untuk klasifikasi multi-kelas. Proses pelatihan dilakukan dengan algoritma *backpropagation* untuk meminimalkan fungsi loss, seperti *cross-entropy*, menggunakan optimisasi *Stochastic Gradient Descent* (SGD) atau Adam. Dataset dibagi menjadi tiga bagian: data latih untuk pelatihan model, data validasi untuk tuning parameter, dan data uji untuk mengevaluasi kinerja. Kinerja MLP diukur menggunakan metrik seperti akurasi, precision, recall, F1-score, dan confusion matrix. Implementasi dilakukan menggunakan framework pembelajaran mesin, seperti TensorFlow atau PyTorch, dengan tuning hiperparameter untuk mendapatkan konfigurasi terbaik.

2.3. Cross-validation

Cross-validation adalah teknik evaluasi model yang digunakan untuk mengukur kinerja algoritma secara lebih akurat dengan meminimalkan risiko overfitting. Dalam penelitian ini, metode k-fold cross-validation diterapkan dengan membagi dataset menjadi *k* lipatan (fold) yang sama besar. Pada setiap iterasi, satu fold digunakan sebagai data uji, sementara *k-1* fold lainnya digunakan sebagai data latih. Proses ini diulang sebanyak *k* kali sehingga setiap fold menjadi data uji satu kali. Rata-rata dari hasil evaluasi pada semua iterasi dihitung untuk menghasilkan metrik performa akhir, seperti akurasi, precision, recall, dan F1-score. Teknik ini memastikan bahwa model diuji pada seluruh data, memberikan evaluasi yang lebih stabil dan generalisasi yang lebih baik. Dalam penelitian ini, nilai *k* dipilih berdasarkan ukuran dataset, dengan nilai umum seperti 5 atau 10. Cross-validation diimplementasikan menggunakan framework pembelajaran mesin seperti *scikit-learn*, yang menyediakan fungsi bawaan untuk mengotomatisasi pembagian dataset dan evaluasi model.

3. Hasil dan Pembahasan

Pengujian dilakukan dengan menggunakan *K-Fold validation*, di mana hasil setiap *fold* akan dibentuk tabel *confusion matrix* dan parameter nilai *k* yaitu *K* = 10.

Tabel 1. Hasil *Confusion Matrix* menggunakan ciri MFCC

Kelas	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	45	0	0	5
D	3	0	0	47	0	0
E	0	0	0	1	49	0
F	0	0	6	0	0	44

Sumber data yang digunakan berasal dari data suara kata yang diambil dari 5 orang penutur. Data suara terdiri dari 6 kelas data yakni A=kakak, B=tutup, C=bibit, D=papan, E=duduk, F=gigit. Setiap orang diminta mengucapkan tiap kelas kata sebanyak 10 kali ujaran, sehingga keseluruhan data adalah 5 orang x 10 ujaran x 6 kelas = 300 ujaran. Data suara kata tersebut kemudian disegmentasi dengan panjang 480 ms. Berdasarkan *confusion matrix* pada Tabel 1, terlihat bahwa akurasi yang dicapai dalam mengklasifikasi isyarat suara tutur menggunakan ciri MFCC untuk 300 data adalah $((48+50+45+47+49+44)/300) \times 100\% = 94,33\%$. *Confusion matrix* memberikan informasi lebih detail tentang performa model untuk tiap kelas. Kelas A: Sebagian besar terklasifikasi dengan benar, namun terdapat 2 kesalahan klasifikasi ke kelas D. Kelas B: Klasifikasi sempurna tanpa kesalahan (50/50). Kelas C: Beberapa kesalahan terjadi, dengan 5 sampel salah diklasifikasikan sebagai kelas F. Kelas D: Terdapat 3 kesalahan klasifikasi ke kelas A. Kelas E: Hanya 1 sampel salah diklasifikasikan ke kelas D. Kelas F: Ada 6 kesalahan klasifikasi ke kelas C. Kelas C dan F akurasi lebih rendah dibandingkan kelas lain karena kata bibit dan gigit memiliki suara dan bentuk isyarat yang agak mirip sehingga menimbulkan ambiguitas pada saat proses klasifikasi. Akan tetapi Penelitian ini mengungkapkan bahwa metode MFCC merupakan teknik yang sangat efektif untuk ekstraksi fitur pada isyarat suara tutur kata yang mengandung konsonan hambat di dalamnya dengan hasil akurasi yang tinggi. MFCC mampu merepresentasikan karakteristik akustik isyarat dengan baik melalui transformasi frekuensi yang menyesuaikan dengan persepsi pendengaran manusia. Proses ekstraksi fitur dalam MFCC mencakup tahapan seperti pre-emphasis, pembagian frame, windowing, transformasi Fourier, dan mel-frequency wrapping, menghasilkan vektor fitur yang ringkas dan informatif. Berdasarkan pengujian, metode MFCC menunjukkan akurasi klasifikasi yang tinggi, mencapai 94,33% untuk klasifikasi 300 data isyarat

suara tutur, terutama pada isyarat dengan tingkat *noise* rendah. Akurasi ini cenderung meningkat dengan panjang data dan ukuran frame sampel yang lebih besar. Keunggulan utama MFCC adalah kemampuannya untuk menangkap detail frekuensi penting dalam isyarat suara, sehingga ideal untuk aplikasi seperti pengenalan suara, analisis emosi, dan verifikasi pembicara. Metode ini berpotensi diterapkan dalam berbagai aplikasi, termasuk pengenalan suara, asisten virtual, dan analisis emosi berbasis suara. Meskipun demikian, performa sistem dipengaruhi oleh kualitas dataset dan tingkat gangguan akustik. Penelitian ini juga menegaskan bahwa mengombinasikan MFCC dengan teknik klasifikasi yang sesuai, seperti metode berbasis probabilistik, dapat meningkatkan kinerja sistem secara keseluruhan.

4. Conclusion

Penelitian ini menunjukkan bahwa metode *Mel-Frequency Cepstral Coefficients* (MFCC) adalah teknik yang efektif untuk ekstraksi ciri dalam isyarat suara tutur. Metode MFCC menunjukkan kemampuan yang efektif dalam merepresentasikan karakteristik akustik isyarat suara melalui proses transformasi frekuensi yang sesuai dengan persepsi pendengaran manusia. Proses ekstraksi meliputi tahapan seperti *pre-emphasis*, *frame blocking*, *windowing*, transformasi Fourier, dan *mel-frequency wrapping*, menghasilkan vektor fitur yang kompak dan informatif. Hasil pengujian menunjukkan bahwa metode MFCC mampu menghasilkan tingkat akurasi klasifikasi yang tinggi, terutama pada kondisi isyarat dengan *noise* rendah. Klasifikasi 300 data isyarat suara tutur menggunakan metode ekstraksi ciri MFCC mencapai akurasi 94,33%. Pada MFCC, akurasi klasifikasi cenderung meningkat seiring dengan bertambahnya panjang data dan ukuran frame sampel. Keunggulan metode ini terletak pada kemampuannya untuk menangkap detail frekuensi penting dari isyarat suara, sehingga cocok untuk berbagai aplikasi seperti pengenalan suara, analisis emosi, dan verifikasi pembicara. Namun, performa sistem dapat dipengaruhi oleh kualitas dataset dan tingkat gangguan akustik. Penelitian ini juga menegaskan bahwa penggunaan MFCC dalam kombinasi dengan teknik klasifikasi yang tepat, seperti metode berbasis probabilistik, dapat meningkatkan kinerja sistem secara keseluruhan. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi penggabungan MFCC dengan fitur tambahan, seperti *delta* dan *delta-delta*, serta penerapan teknik pengurangan *noise* untuk meningkatkan akurasi dalam lingkungan yang lebih kompleks. Disarankan juga untuk menambah jumlah rekaman dan variasi pembicara dalam tahap pelatihan maupun pengujian, agar dapat mengurangi variabilitas pembicara dalam pengenalan, sesuai dengan standar yang telah ditetapkan dalam penelitian sebelumnya.

Referensi

- [1] H. R. Paleva and B. H. Prasetyo, "Penerapan Short Time Fourier Transform pada MFCC untuk Sistem Pengenalan Ucapan Tingkat Stres," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, pp. 1–10, 2024.
- [2] G. Ajinurseto, L. O. Bakrim, and N. Islamuddin, "Penerapan Metode Mel Frequency Cepstral Coefficients pada Sistem Pengenalan Suara Berbasis Desktop," *INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi*, vol. 25, no. 1, pp. 11–20, 2023.
- [3] S. Santoso, R. Hartayu, C. Anam, and Dimas Abdul Aziz, "Simulasi Simulasi Ekstraksi Fitur Suara menggunakan Mel-Frequency Cepstrum Coefficient," *Jurnal Sains dan Informatika*, vol. 8, no. 1, pp. 80–87, 2022.
- [4] I. P. B. C. Buana, I. G. A. Wibawa, A. Muliantara, M. A. Raharja, I. P. G. H. Suputra, and L. G. Astuti, "Deteksi Suara Paru-Paru Menggunakan MFCC (Mel Frequency Cepstral Coefficient) dan M-KNN (Modified K-Nearest Neighbor)," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 11, no. 4, p. 845, 2022.
- [5] A. A. Sasilo, R. A. Saputra, and I. P. Ningrum, "Sistem Pengenalan Suara Dengan Metode Mel Frequency Cepstral Coefficients Dan Gaussian Mixture Model," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 203–210, 2022.
- [6] P. Tridarma and S. N. Endah, "Pengenalan Ucapan Bahasa Indonesia Menggunakan MFCC dan Recurrent Neural Network," *Jurnal Masyarakat Informatika*, vol. 11, no. 2, pp. 36–44, 2020.
- [7] R. Hidayat and A. Winursito, "Improving accuracy of isolated word recognition system by using syllable number characteristics," *International Journal of Technology*, vol. 11, no. 2, pp. 411–421, 2020.
- [8] A. Anggoro, S. Herdjunto, and R. Hidayat, "MFCC dan KNN untuk Pengenalan Suara Artikulasi P," *Avitec*, vol. 2, no. 1, pp. 13–19, 2020.
- [9] D. Prabakaran and S. Sriuppili, "Speech processing: MFCC based feature extraction techniques - An investigation," *Journal of Physics: Conference Series*, vol. 1717, no. 1, 2021.

- [10] I. S. Permana, Y. Indrawaty, and A. Zulkarnain, "Implementasi Metode Mfcc Dan Dtw Untuk Pengenalan Jenis Suara Pria Dan Wanita," *MIND Journal*, vol. 3, no. 1, pp. 61–76, 2019.