NARATIF : Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika
Vol. 06 No. 01 Juni 2024
P-ISSN: 2656-7377 || E-ISSN: 2714-8467

PERBANDINGAN KUALITAS PENGENALAN SUARA UNTUK EKSTRAKSI FITUR MENGGUNAKAN MFCC DAN *SPECTRAL*

Ricky Aurelius N. Diaz[1]*, Ni Luh Gede Pivin Suwirmayanti[2], Komang Budiarta[3]
Program Studi Sistem Komputer[1,2], Program Studi Sistem Informasi[3]
Institut Teknologi dan Bisnis STIKOM Bali,Indonesia
Denpasar, Indonesia
e-mail: [1]ricky@stikom-bali.ac.id, [2]pivin@stikom-bali.ac.id, [3]komang_budiarta@stikom-bali.ac.id

Abstrak

Tahapan awal dalam pengenalan suara adalah tahap ekstraksi fitur, dimana penerapan metode sangatlah dapat berdampak signifikan terhadap kualitas pengenalan suara, sehingga perlu dilakukan pemilihan metode yang tepat. Metode ekstraksi fitur untuk pengenalan suara diantaranya *Mel-Frequency Cepstral Coefficients* (MFCC) dan representasi spektral. MFCC telah menjadi standar dalam berbagai aplikasi pengenalan suara karena kemampuannya dalam menangkap karakteristik penting dari suara manusia. Sementara itu, representasi spektral memiliki pendekatan yang lebih sederhana dengan hanya menganalisis amplitudo spektrum suara tanpa mempertimbangkan informasi cepstral. Penelitian ini menggunakan ekstraksi fitur yang dilakukan dengan mengimplementasikan kedua metode, yaitu MFCC dan representasi spektral, pada setiap sampel suara dalam *dataset*. Selanjutnya, dilakukan pemrosesan data menggunakan algoritma pengenalan pola seperti *K-Nearest Neighbors* (K-NN) untuk mengklasifikasikan suara pada kedua kelompok fitur. Hasil penelitian ini diperoleh hasil bahwa MFCC memiliki keunggulan dalam proses identifikasi suara berbasis gender dengan akurasi tertinggi 84,18 untuk data training dan 74,71 untuk data *testing* dimana kedua hasil ini berasal dari kelompok data yang sama yaitu pembagian 50% data uji dan 50% data *training*. Hasil eksperimen menunjukkan bahwa penggunaan MFCC cenderung menghasilkan hasil pengenalan suara yang lebih baik dibandingkan dengan representasi spektral. Hal ini disebabkan oleh kemampuan MFCC dalam menangkap informasi frekuensi dan temporal dari suara manusia.
Kata kunci**:** Perbandingan, Ekstraksi Fitur, MFCC, *Spectral*, KNN.

Abstract

Voice recognition is one of the important areas in audio signal processing and voice-based applications. The initial stage is the feature extraction stage, where the application of the method can have a significant impact on the quality of voice recognition, so it is necessary to choose the right method. One of the feature extraction methods for speech recognition is the Mel-Frequency Cepstral Coefficient (MFCC) and spectral representation. The MFCC feature extraction method has become a standard in various speech recognition applications due to its ability to capture important characteristics of the human voice. Meanwhile, spectral representation has a simpler approach by only analyzing the amplitude of the sound spectrum without considering cepstral information. This research uses a voice dataset that includes a wide variety of speech and accents. Feature extraction is carried out by implementing both methods, namely MFCC and spectral representation, on each sound sample in the dataset. Next, data processing is carried out using a pattern recognition algorithm such as K-Nearest Neighbors (K-NN) to classify sounds in both feature groups. The results of this research show that MFCC has superiority in the gender-based voice identification process with the highest accuracy of 84.18 for training data and 74.71 for testing data where these two results come from the same data group, namely a division of 50% test data and 50% data training. Experimental results show that the use of MFCC tends to produce better speech recognition results compared to spectral representation. This is due to the MFCC's ability to capture frequency and temporal information from human voices.
Keywords: comparison; feature extraction; MFCC; Spectral ; KNN.

I. PENDAHULUAN

Potensi berkembangnya pengenalan bahasa lisan ini terbantu dengan banyaknya data suara yang tersedia di *internet* saat ini sehingga dapat dimanfaatkan untuk keperluan *machine learning* dan pembentukan *dataset*. Dengan tersedianya data saat ini, dimana bentuknya adalah sinyal digital, maka tersedia kemungkinan untuk adanya penelitian dalam bidang pengenalan bahasa lisan ini dengan melihat kemampuan metode dalam proses ekstraksi fitur, kemampuan algortima dalam mengelompokkan jenis bahasa atau secara khusus kemampuan dalam proses identifikasi. Dalam upaya untuk meningkatkan kualitas pengenalan suara, pemilihan metode ekstraksi fitur yang tepat memiliki peran yang sangat penting. *Mel-Frequency Cepstral Coefficients* (MFCC) dan representasi spektral adalah dua metode umum yang digunakan untuk ekstraksi fitur dalam pengenalan suara. MFCC telah terbukti menjadi pilihan yang efektif dalam banyak kasus, karena mampu menangkap karakteristik frekuensi dan temporal dari sinyal suara dengan baik. Di sisi lain, representasi spektral adalah pendekatan yang lebih sederhana, berfokus pada analisis amplitudo spektrum suara tanpa mempertimbangkan aspek-aspek *cepstral*.

Melihat kondisi tersebut, maka dalam penelitian ini, proses pengenalan bahasa lisan akan dimulai dengan memanfaatkan beberapa teknik ekstraksi fitur sebagai fondasi awal proses pengenalan, dimana hasilnya akan dibandingkan dari sisi akurasi menggunakan algortima klasifikasi umum yaitu KNN. Data yang akan digunakan dalam penelitian ini adalah *dataset* yang berasal dari *VoxCeleb* adalah kumpulan data audio-visual yang terdiri dari klip pendek ucapan manusia, diambil dari video wawancara yang diunggah ke *YouTube*. *VoxCeleb* berisi pidato dari pembicara yang mencakup berbagai etnis, aksen, profesi, dan usia yang berbeda. Hasil penelitian ini menyertakan perbandingan antara penggunaan MFCC dan representasi spektral dalam konteks ekstraksi fitur untuk pengenalan suara. Tujuan utama dari penelitian ini adalah untuk menganalisis perbedaan kualitas pengenalan suara yang dihasilkan oleh kedua metode

tersebut. Selain itu, penelitian ini juga akan mencakup penggunaan algoritma klasifikasi seperti *K-Nearest Neighbors* (K-NN) dalam mengolah data ekstraksi fitur, dengan tujuan untuk memahami bagaimana perbedaan ekstraksi fitur dapat mempengaruhi performa klasifikasi. Dalam penelitian sebelumnya, KNN digunakan untuk klasifikasi tangisan bayi, dimana akurasi terbaik pada proses klasifikasi menggunakan data sampling *Percentage Rate* yaitu 76% dengan nilai K yang digunakan adalah 9, sedangkan akurasi terbaik pada proses klasifikasi menggunakan data sampling *Leave One Out* yaitu 42% dengan nilai K yang digunakan adalah 5[1]. Hasil penelitian ini dapat memberikan informasi dalam pemilihan metode ekstraksi fitur yang tepat dalam aplikasi pengenalan suara.

II. TINJAUAN PUSTAKA

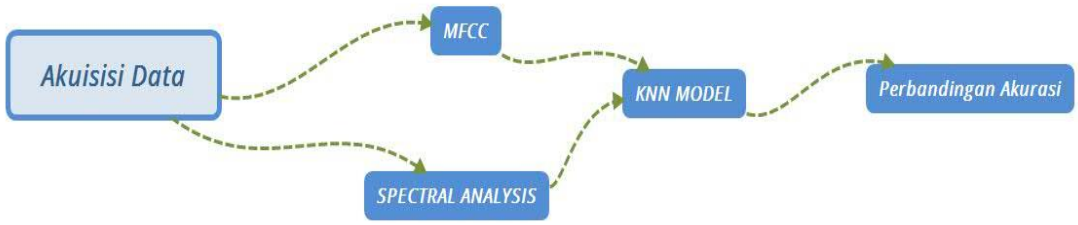
Mel-Frequency Cepstral Coefficients (MFCC) adalah fitur audio yang umum digunakan dalam pengenalan suara, karena kemampuannya dalam menggambarkan variasi sinyal dalam frekuensi rendah. MFCC menggambarkan energi *cepstrum* dalam skala non-linier dan dianggap sebagai fitur akustik yang paling mendekati bagaimana cara manusia mendengarkan sinyal ucapan [2]. Pengenalan pola atau identifikasi dikenal berbagai algoritma data mining yang dapat digunakan untuk klasifikasi seperti *Support Vector Machine* (SVM), *Hidden Markov Model* (HMM), *Gaussian Mixture Model* (GMM), *Random Forest*, *Decision Tree* dan berbagai algoritma lainnya yang dapat digunakan baik secara tunggal maupun diuji secara *hybrid*.

Sebelumnya, penelitian dengan judul *Comparative performance analysis for speech digit recognition based on MFCC and vector quantization* yang dilakukan pada tahun 2021, melakukan analisis perbandingan analisis *cepstral* dengan MFCC menggunakan teknik pencocokan fitur kuantisasi vektor. Semua digit ucapan dari nol ucapan hingga sembilan digit data ucapan telah dikumpulkan untuk 15 subjek dalam tiga sesi berbeda. Penelitian ini menunjukkan bahwa teknik ekstraksi fitur MFCC memberikan kinerja yang lebih baik dibandingkan fitur *cepstral* untuk data ucapan digit lisan[3]. Pada penelitian lain dalam pengenalan suara tembang sekar alit, diperoleh hasil bahwa algoritma MFCC memberikan akurasi yang cukup baik yaitu sebesar 76,6% [4]. Beberapa penelitian lain juga menunjukkan bahwa MFCC cocok untuk digunakan dalam ekstraksi ciri pola pengucapan baik bahasa Inggris maupun pola pengucapan dalam bahasa jawa [5], [6]. Dari sisi algoritma yang digunakan, beberapa algoritma klasifikasi sering digunakan untuk dipasangkan dengan MFCC maupun *spectral* seperti SVM, HMM, GMM, serta *Random Forest*, dan memiliki hasil akurasi yang cukup baik. [7], [8], [9], [10], [11]

III. HASIL DAN PEMBAHASAN

1. Metode Penelitian

Proses perbandingan dimulai dengan tahapan pengumpulan data berupa *speech audio* dan dilanjutkan dengan ekstraksi fitur menggunakan *spectral* analysis dan MFCC. Hasil ekstraksi fitur kemudian menjadi dasar untuk membentuk data latih dan data uji. Proses latih dan uji data ini dilakukan dengan KNN Model dan selanjutnya melihat hasil ekstraksi terbaik antara fitur *Spectral* atau MFCC yang digunakan. Metode penelitian dalam usulan ini dijabarkan pada gambar berikut:



Gambar 1. Metode Penelitian

2. Akuisisi Data

Proses akuisisi data merupakan proses persiapan data yang akan digunakan untuk *training* dan *testing*, dimana data diperoleh dari *VoxCeleb* dengan menggunakan kurang lebih 7000 audio ucapan unik yang terbagi dalam 3683 suara pria dan 2312 suara wanita.

3. Ekstraksi Fitur

Pada tahap ini, *file* audio yang telah dimiliki akan diproses untuk mendapatkan ciri-ciri khusus yang disebut dengan ekstraksi fitur. Proses ekstraksi dimulai dengan mengubah sinyal *file* audio yang menjadi domain frekuensi menggunakan metode *Discrete Cosinus Transform*. Selanjutnya, sinyal tadi akan diolah dengan menggunakan *spectral* analisis untuk mendapatkan fitur khusus dari masing-masing sinyal. MFCC sebagai salah satu metode ekstraksi fitur juga akan menjadi dasar dalam proses perbandingan pada tahap selanjutnya bersama dengan hasil *spectral* analisis. Berikut ini adalah gambaran hasil ekstraksi fitur untuk *Spectral* Fitur dan MFCC.

TABEL I.
FITUR *SPECTRAL*

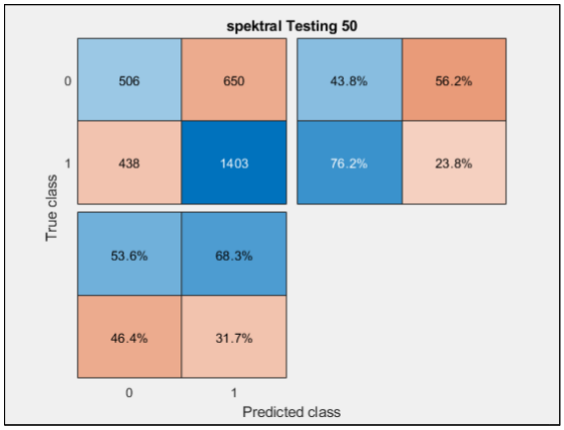
Nama *File*	Centroid	Flatness	Kurtosis	Rolloff	Skewness
0.m4a'	3658	0.64557	287.5	6422.6	14.913
'1.m4a'	5810.7	0.68542	278.46	6891.2	15.473
'10.m4a'	3119.1	0.68528	233.82	6165	13.886
'100.m4a'	5487.6	0.74841	227.21	7351.9	13.758
'1000.m4a'	2093.9	0.68396	158.15	6524.2	10.473
….	….	….	….	….	….

TABEL II.
FITUR MFCC

Nama *File*	Attr 1	Attr 2	Attr 3	Attr 4	..Attr14
0.m4a'	3.3129	-9.0764	4.1021	3.0073	1.4125
'1.m4a'	1.333	-11.755	5.6541	2.5192	0.79279
'10.m4a'	1.0513	-12.347	5.5701	2.1744	0.47532
'100.m4a'	3.9894	-7.6628	4.0195	2.3208	1.0406
'1000.m4a'	2.6906	-10.062	3.5079	1.8013	0.84764
….	….	….	….	….	….

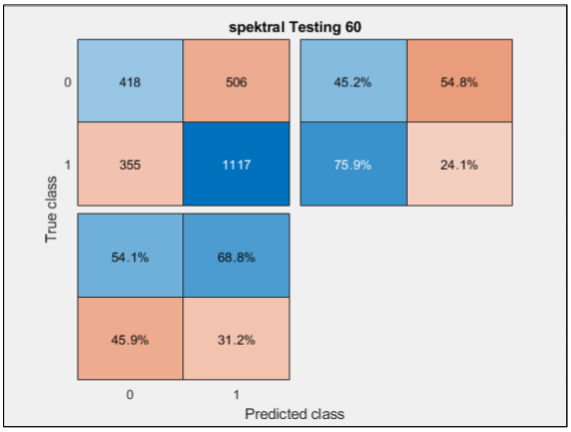
4. Identifikasi dengan KNN

Hasil dari tahap ekstraski fitur selanjutnya digunakan sebagai dasar untuk proses identifikasi menggunakan KNN Model. Kombinasi pengujian yang digunakan menggunakan tiga jenis yaitu perbandingan data *training* dengan data *testing* sejumlah 50%:50%, 60%:40% dan 70%:30% (%data *training* : %data *testing*). Berikut ini adalah hasil proses identifikasi dengan menggunakan KNN menggunakan *Spectral* Fitur :



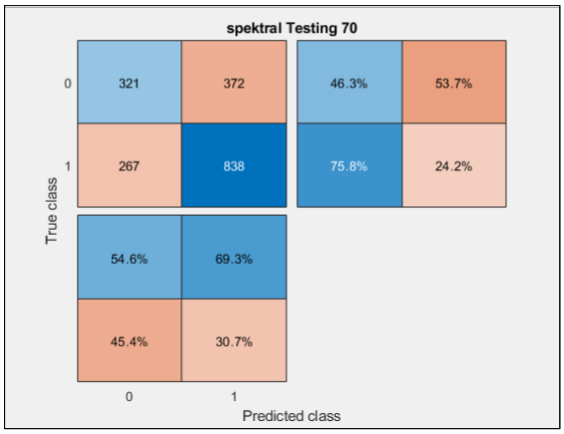
Gambar 2. Pengujian *Spectral* dengan 50% Data

Pada Gambar 2, terlihat hasil pengujian terhadap spektral fitur dengan menggunakan ukuran data 50%:50% (%data *training* : %data *testing*), dimana pengujian menunjukkan hasil bahwa 506 data merupakan *True Negative*, 650 merupakan *False Positive*, 438 merupakan *False Negative* dan 1403 data merupakan *True Positive* dari total data 2997 data, dan berturut-turut nilai *recall* untuk *True Positive* adalah 76,2% serta *Precission* sebesar 68,3%.



Gambar 3. Pengujian *Spectral* dengan 60% Data

Gambar 3 menunjukkan pengujian terhadap spektral fitur dengan menggunakan ukuran data 60%:40% (%data *training* : %data *testing*), dimana pengujian menunjukkan hasil bahwa 418 data merupakan *True Negative*, 506 merupakan *False Positive*, 355 merupakan *False Negative* dan 1117 merupakan *True Positive* dari total 2396 data, dan menghasilkan nilai *recall* untuk *True Positive* adalah 75,9% serta *Precission* sebesar 68,8%.



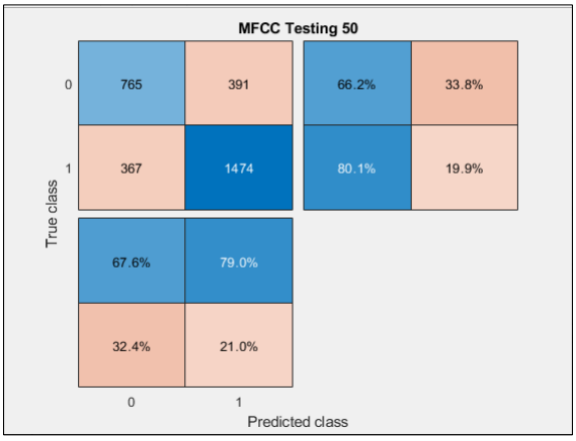
Gambar 4. Pengujian *Spectral* dengan 70% Data

Pada Gambar 4, dilakukan pengujian terhadap spektral fitur dengan menggunakan ukuran data 70%:30% (%data *training* : %data *testing*), dimana pengujian menunjukkan hasil bahwa 321 data merupakan *True Negative*, 372 merupakan *False Positive*, 267 merupakan *False Negative* dan 838 merupakan *True Positive* dari total data 1798 data, dan berturut-turut nilai *recall* untuk *True Positive* adalah 75,3% serta *Precission* sebesar 69,3%. Dari hasil identifikasi menggunakan KNN yang telah ditunjukkan oleh *confusion matrix* pada gambar 2, gambar 3 dan gambar 4, diperoleh hasil keseluruhan akurasi untuk *Spectral* Fitur sebagai berikut :

TABEL III.
HASIL PENGUJIAN *SPECTRAL* FITUR

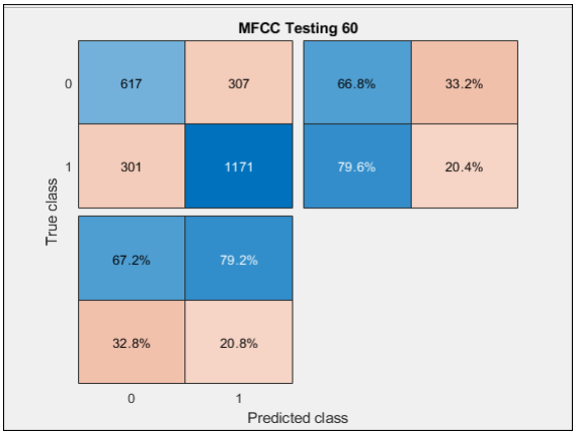
Data	**Hasil *Training***	**Hasil *Testing***
50%	74.67	63.7
60%	75.95	64.07
70%	74.92	64.46

Selanjutnya percobaan dilakukan dengan menggunakan fitur MFCC, dimana berikut ini adalah hasil proses identifikasi dengan menggunakan MFCC :



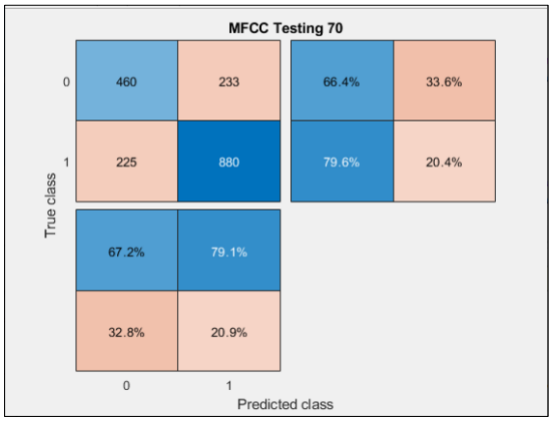
Gambar 5. Pengujian MFCC dengan 50% Data

Gambar 5 menunjukkan hasil pengujian terhadap fitur dari MFCC dengan menggunakan ukuran data 50%:50% (%data *training* : %data *testing*), dimana pengujian menunjukkan hasil bahwa 765 data merupakan *True Negative*, 391 data merupakan *False Positive*, 367 merupakan *False Negative* dan 1474 data merupakan *True Positive* dari total data 2997 data, dan berturut-turut nilai *recall* untuk *True Positive* adalah 80,1% serta *Precission* sebesar 79%.



Gambar 6. Pengujian MFCC dengan 60% Data

Pengujian dilanjutkan dengan menggunakan ukuran data 60%:40% (%data *training* : %data *testing*), seperti pada gambar 6 diatas, dimana hasil pengujian menunjukkan bahwa 617 data merupakan *True Negative*, 307 data merupakan *False Positive*, 301 data merupakan *False Negative* dan 1171 data merupakan *True Positive* dari total data 2396 data, dengan nilai *recall* untuk *True Positive* adalah 79,6% serta *Precission* sebesar 79,2%.



Gambar 7. Pengujian MFCC dengan 70% Data

Terakhir, pengujian dilakukan dengan menggunakan ukuran data 70%:30% (%data *training* : %data *testing*), seperti yang terlihat pada gambar 7, dimana hasil pengujian menunjukkan bahwa 460 data merupakan *True Negative*, 233 data merupakan *False Positive*, 225 data merupakan *False Negative* dan 880 data merupakan *True Positive* dari total data 1798 data, dengan nilai *recall* untuk *True Positive* adalah 79,6% serta *Precission* sebesar 79,1%. Secara keseluruhan, hasil identifikasi menggunakan KNN untuk fitur MFCC adalah sebagai berikut :

TABEL IV. HASIL PENGUJIAN FITUR MFCC

Data	Hasil *Training*	Hasil *Testing*
50%	84.18	74.71
60%	83.49	74.62
70%	83.81	74.53

Seperti yang ditunjukkan oleh Tabel IV, hasil pengujian fitur MFCC menunjukkan akurasi keseluruhan training dan *testing*, dimana akurasi tertinggi untuk *training* adalah sebesar 84,18% pada ukuran data 50%, dan akurasi *testing* paling besar adalah 74,71% juga pada ukuran data 50%.

IV. KESIMPULAN

Dari hasil pengujian yang telah dilakukan, diperoleh hasil bahwa MFCC memiliki keunggulan dalam proses identifikasi suara berbasis gender dibandingkan dengan *Spectral* Fitur, dimana MFCC memiliki akurasi tertinggi 84,18 untuk data *training* dan 74,71 untuk data *testing* dimana kedua hasil ini berasal dari kelompok data yang sama yaitu pembagian 50% data uji dan 50% data *training*. Adapun saran yang dapat dijadikan sebagai dasar perbaikan dari penelitian ini adalah menggunakan fitur *spectral* yang lebih luas untuk percobaan meningkatkan nilai akurasi, serta dapat memanfaatkan model atau algoritma klasifikasi yang berbeda untuk proses perbandingan atau peningkatan akurasi.

REFERENSI

[1] A. S. Prayogi, M. Rizqi, and T. M. Fahrudin, "Klasifikasi Suara Tangisan Bayi Berdasarkan Prosodic Features Menggunakan Metode Moments of Distribution dan K-Nearest Neighbours," *Teknika*, vol. 8, no. 2, pp. 119–125, Oct. 2019, doi: 10.34148/teknika.v8i2.206.

[2] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi, and A. Othmani, "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech," *Biomed Signal Process Control*, vol. 71, Jan. 2022, doi: 10.1016/j.bspc.2021.103107.

[3] D. R. KS, R. MD, and S. G, "Comparative performance analysis for speech digit recognition based on MFCC and vector quantization," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 513–519, Nov. 2021, doi: 10.1016/j.gltp.2021.08.013.

[4] M. Agung Raharja *et al.*, "IMPLEMENTASI METODE MEL-FREQUENCY CEPSTRAL COEFFICIENT DAN DTW PADA APLIKASI PENGENALAN SUARA TEMBANG SEKAR ALIT", [Online]. Available: https://s.id/jurnalresistor

[5] I. K. Almanfaluti and J. P. Sugiono, "Identifikasi Pola Suara Pada Bahasa Jawa Meggunakan Mel Frequency Cepstral Coefficients (MFCC)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 1, p. 22, Jan. 2020, doi: 10.30865/mib.v4i1.1793.

[6] M. Azhar and H. F. Pardede, "Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 439, Apr. 2021, doi: 10.30865/mib.v5i2.2754.

[7] R. B. Handoko and S. Suyanto, "Klasifikasi Gender Berdasarkan Suara Menggunakan Support Vector Machine," *Indonesian Journal on Computing (Indo-JC)*, vol. 4, no. 1, p. 9, Mar. 2019, doi: 10.21108/indojc.2019.4.1.244.

[8] I. S. El Bashart and T. Pangaribowo, "Aplikasi Pengenalan Suara untuk Pengamanan Software Komputer Menggunakan Metode MFCC(Mel Frequency Cepstrum Coefficients) dan HMM (Hidden Markov Model)," *Jurnal Teknologi Elektro*, vol. 11, no. 1, p. 39, Feb. 2020, doi: 10.22441/jte.2020.v11i1.006.

[9] A. A. Sundawa, A. G. Putrada, and N. A. Suwastika, "Implementasi dan Analisis Simulasi Deteksi Emosi Melalui Pengenalan Suara Menggunakan Mel-Frequency Cepstrum Coefficient dan Hidden Markov Model Berbasis IOT."

[10] K. K. Rekayasa, F. Dharma Adhinata, D. Putra Rakhmadani, A. Jala, and T. Segara, "Terbit online pada laman web jurnal: http://journal.ittelkom-pwt.ac.id/index.php/dinda JURNAL DINDA Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM," 2021. [Online]. Available: https://research.google.com/audioset.

[11] A. Septiani and A. Rizal, "Klasifikasi Suara Paru Normal dan Abnormal dengan Menggunakan Discrete Wavelet Transform dan Support Vector Machine."