Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 21, No. 3, Oktober 2024, pp.260-271
ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339

Implementation of Mel-Frequency Cepstral Coefficient as Feature Extraction Method On Speech Audio Data

Implementasi Mel-Frequency Cepstral Coefficient Sebagai Metode Ekstraksi Ciri Pada Data Audio Suara Ucapan

Andre Julio Sumurung Marbun[1], Heriyanto[2], Frans Richard Kodong[3]

[1,2,3]Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

[1]123170079@student.upnyk.ac.id, [2*]heriyanto@upnyk.ac.id, [3]frans.richard@upnyk.ac.id
*: Penulis korenspondensi (corresponding author)

Informasi Artikel
Received: December 2023
Revised: March 2024
Accepted: August 2024
Published: October 2024

Abstract

Purpose: To determine whether the combination of MFCC as a feature extraction and SVM as a classification method can be used for speaker emotion classification based on speech sounds on the RAVDESS dataset using software assistance.
Design/methodology/approach: This research uses RAVDESS research dataset as training data and test data. The methods used in this research are MFCC as a method of character extraction on audio data and SVM as a method of emotion classification.
Findings/result: The best accuracy in males of 85.71% was obtained with a combination of filter parameter pre emphasize of 0.95, frame length of 0.023 ms, overlap of adjacent frames of 40%, number of mel filters in the melscaled filterbank process of 24 mel, number of cepstral coefficient of 24 coefficient and the value of 'C' in SVM of 0.01. The best accuracy in women of 92.21% was obtained with a combination of filter parameters pre-emphasize of 0.95, frame length of 0.023 ms, overlap of adjacent frames of 40%, the number of mel filters in the mel-scaled filterbank process of 24 mel, and the number of cepstral coefficient of 13 coefficient and 'C' value in SVM of 0.01. From the two test results of tuning parameters between men and women, there are similar parameter values in all test parameters, except for the number of cepstral coefficients. The number of cepstral coefficient in men is 24 coefficient while the number of cepstral coefficient in women is 13 coefficient.
Originality/value/state of the art: The combination of feature extraction method and classification method is based on previous studies. Furthermore, the method is used to determine the speaker's emotion on the RAVDESS research dataset.

Keywords: Mel-Frequency Cepstral Coefficient; Support Vector Machine; Classification Emotion; Signal processing
Kata kunci: Mel-Frequency Cepstral Coefficient; Support Vector Machine; Klasifikasi emosi; Pemrosesan sinyal

Abstrak

Perancangan/metode/pendekatan: Penelitian ini menggunakan dataset penelitian RAVDESS sebagai data latih dan data uji. Metode yang digunakan pada penelitian ini yaitu MFCC sebagai metode ekstraksi ciri pada data audio dan SVM sebagai metode klasifikasi emosi.
Hasil: Akurasi terbaik pada laki-laki sebesar 85,71% diperoleh dengan kombinasi parameter filter pre-emphasize sebesar 0.95, panjang frame sebesar 0.023 ms, overlap frame yang bersebelahan sebesar 40%, jumlah filter mel pada proses mel-scaled filterbank sebesar 24 mel, jumlah cepstral coefficient sebesar 24 coefficient dan nilai 'C' pada SVM sebesar 0,01. Akurasi terbaik pada wanita sebesar 92,21% diperoleh dengan kombinasi parameter filter preemphasize sebesar 0.95, panjang frame sebesar 0.023 ms, overlap frame yang bersebelahan sebesar 40%, jumlah filter mel pada proses mel-scaled filterbank sebesar 24 mel, dan jumlah cepstral coefficient sebesar 13 coefficient dan nilai 'C' pada SVM sebesar 0,01. Dari kedua hasil pengujian parameter tuning antara laki-laki dan wanita terdapat kesamaan nilai parameter pada semua parameter pengujian, kecuali pada jumlah cepstral coefficient. Jumlah cepstral coefficient pada laki-laki sebanyak 24 coefficient sedangkan jumlah cepstral coefficient pada wanita sebanyak 13 coefficient.
Keaslian/state of the art: Gabungan antara metode ekstraksi ciri dan metode klasifikasi diambil berdasarkan penelitianpenelitian sebelumnya. Selanjutnya metode tersebut digunakan untuk mengetahui emosi pembicara pada dataset penelitian RAVDESS.

1. Pendahuluan

Suara tidak dapat langsung diolah oleh mesin tanpa adanya proses ekstraksi ciri yang dilakukan terlebih dahulu. Saat ini terdapat begitu banyak pilihan metode ekstraksi ciri yang dapat digunakan, sehingga menentukan metode ekstraksi ciri yang tepat merupakan hal yang tidak mudah. Hal ini juga ditegaskan oleh Helmiyah (2021) yang menyatakan bahwa pemilihan metode ekstraksi ciri pada suara merupakan tahap yang penting dan berpengaruh terhadap hasil akhir pembelajaran mesin karena banyak raw data yang dapat mempengaruhi dalam menemukan ciri suara [1]. Selain itu Ancilin dan Milton (2021) sepakat bahwa mengidentifikasi ciri yang tepat pada suara merupakan suatu permasalahan menantang karena ciri yang berhasil diekstrak perlu menekankan informasi yang dibutuhkan karena banyak yang dapat faktor lingkungan dan suara lain dalam suatu informasi [2].

Muljono et al. (2019) memilih dataset dialog film untuk menguji kemampuan MFCC dalam mengekstrak informasi dari data potongan dialog film yang dipengaruhi oleh noise backsound [3]. Noise backsound pada dialog film mempengaruhi akurasi akhir, sehingga Muljono (2019) hanya mampu menghasilkan akurasi sebesar 66% dengan metode klasifikasi SVM kernel *linear*. Aini, Santoso, dan Dutono (2021) menggunakan MFCC sebagai metode ekstraksi ciri dan untuk deteksi emosi berdasarkan ucapan dalam bahasa Indonesia karena MFCC memiliki ketahanan terhadap kebisingan dari lingkungan serta definisi emosi yang ambigu dan pembagian yang

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339

tidak jelas pada emosi yang berbeda [4]. Al Dujaili et al. (2021), menyatakan bahwa suara manusia terdiri dari beberapa kombinasi informasi seperti maskud harfiah pembicara, budaya pembicara, kondisi psikologikal pembicara dan kondisi emosi pembicara [5]. Hal-hal tersebut dapat menjadi variabel yang mempengaruhi kinerja model klasifikasi emosi sehingga diperlukan pertimbangan yang cermat saat proses pembuatan model untuk mengurangi pengaruh varibel-variabel tersebut. Oleh karena itu Al Dujaili et al. (2021), menggabungkan beberapa metode ekstraksi ciri seperti Fundamental Frequency, energy, ZeroCrossing Rate (ZCR), dan Fourier parameter untuk mendapatkan informasi yang dibutuhkan, kemudian informasi tersebut dimasukkan dan digunakan untuk melatih model klasifikasi emosi.

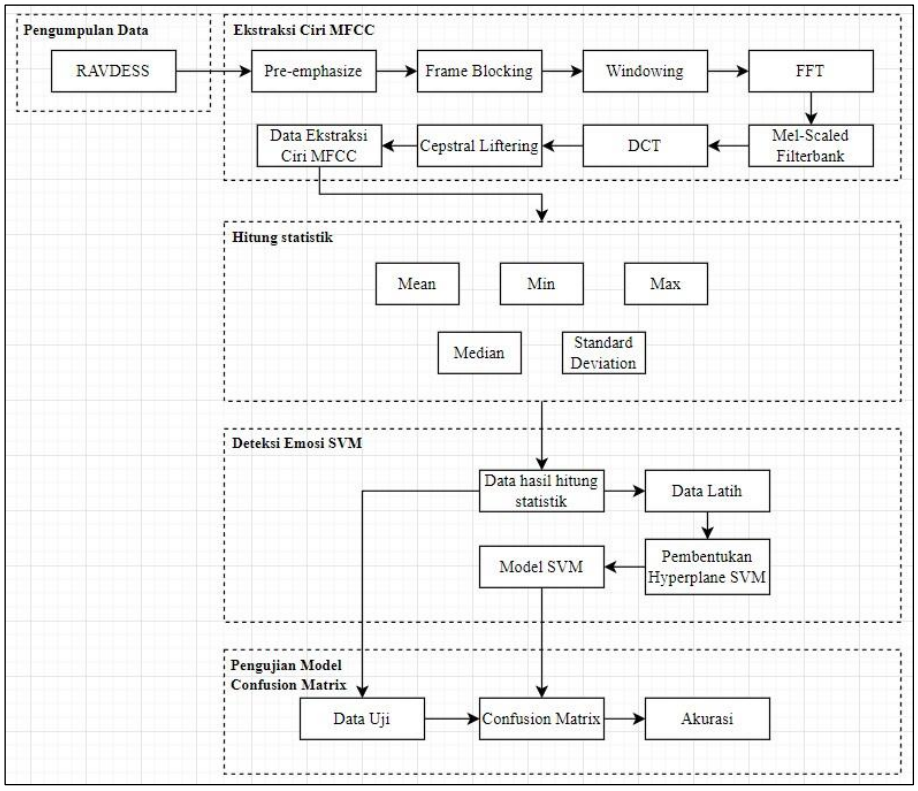
Salah satu metode ekstraksi ciri pada sinyal suara yang kerap kali digunakan yaitu MelFrequency Cepstral Coefficient (MFCC). MFCC merupakan metode ekstraksi ciri pada sinyal suara yang memiliki prinsip kerja seperti telinga manusia [1]. Hal ini menyebabkan MFCC banyak digunakan pada berbagai tugas yang berhubungan dengan pengenalan berdasarkan sinyal suara karena sesuai dengan prinsip kerja telinga manusia banyak suara yang ditangkap oleh telinga.

Penelitian ini akan menggunakan metode MFCC untuk mengekstrak ciri pada sinyal suara. Data hasil ekstraksi ciri kemudian digunakan untuk mengklasifikasikan emosi pembicara. Metode yang digunakan untuk mengklasifikasikan emosi pembicara adalah *Support Vector Machine*. *Support Vector Machine* dipilih karena metode tersebut tidak memiliki banyak parameter yang perlu diatur. Dataset suara yang digunakan pada penelitian ini merupakan dataset penelitian Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).

2. Metode Penelitian

Metodologi penelitian merupakan gambaran terkait alur kerja atau tahapan dalam penelitian. Alur penelitian yang dilakukan ditunjukkan pada Gambar 1. Penelitian dimulai dengan mengunduh data audio RAVDESS dari *website Kaggle*. Data audio kemudian diekstrak cirinya dengan menggunakan metode MFCC. Data hasil ekstraksi ciri MFCC selanjutnya dihitung untuk mendapatkan nilai *mean*, *max, median, min*, dan *standard deviation*. Data hasil hitung statistik kemudian dibagi menjadi data latih dan data uji. Data latih digunakan untuk melatih model SVM dan data uji digunakan untuk menguji model SVM. Model diuji dengan menggunakan metode *confusion matrix.*

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339



Gambar 1. Alur tahapan penelitian

2.2. Pengumpulan Data

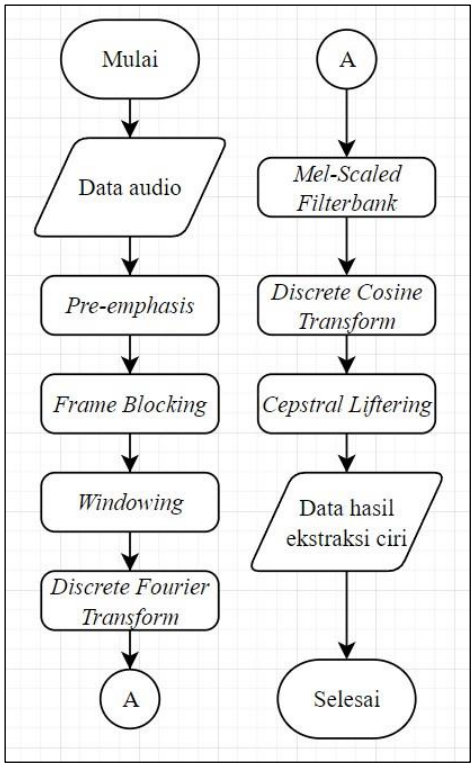
Data yang digunakan pada penelitian ini yaitu data *Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS) yang diunduh dari *website Kaggle*. RAVDESS merupakan dataset audio dan video yang diproduksi oleh Livingstone dan Russo pada tahun 2018. RAVDESS secara spesifik diproduksi untuk memenuhi kebutuhan akan minimnya validitas dataset yang berkaitan dengan emosi manusia. Penelitian ini selanjutnya hanya menggunakan dataset audio yang telah disediakan oleh RAVDESS.

Data audio RAVDESS memiliki format *file wav, sampling rate* 48000 Hz, dan hanya memiliki satu *channel* atau mono. Data audio RAVDESS diucapkan oleh 12 aktor laki-laki dan 12 aktor wanita yang menyuarakan dua pernyataan yang sama dalam bahasa Inggris dengan aksen Amerika Utara yang netral. Terdapat dua jenis tipe intensitas emosi pada data audio RAVDESS, data audio dengan intensitas emosi yang kuat dan intensitas emosi yang normal.

2.3. Mel-Frequency Cepstral Coefficient

Mel-Frequency Cepstral Coefficients (MFCC) memiliki prinsip kerja menyerupai sistem pendengaran manusia. MFCC secara umum terdiri dari beberapa proses pengolahan sinyal yaitu *Pre-emphasis, Frame Blocking, Windowing, Fast Fourier Transform, Mel-Scaled Filterbank,* dan *Discrete Cosine Transform.* Umumnya MFCC berhenti setelah proses DCT, namun beberapa penelitian lain menggunakan *Cepstral Liftering* setelah proses DCT. Putra (2011) dalam penelitiannya menjelaskan bahwa proses *Cepstral Liftering* dapat meningkatkan akurasi pengenalan pola pada sinyal suara [6]. *Flowchart* MFCC dapat dilihat pada Gambar 2.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339



Gambar 2. *Flowchart* MFCC

1. *Pre-emphasis*

   Pre-emphasis adalah proses yang dilakukan untuk meningkatkan kualitas sinyal dengan cara menekan sinyal pada frekuensi tinggi sehingga selaras dengan frekuensi rendah [7]. Proses ini dilakukan karena sinyal sering sekali mengalami gangguan (*noise*), sehingga diperlukan proses untuk mengurangi *noise* pada sinyal [8]. *Pre-emphasis* dilakukan pada seluruh data di dalam sinyal audio. Persamaan *pre-emphasis* sebagai berikut:

   $y(n) = s(n) - \alpha\, s(n-1)$ ..... (2.1)

   Keterangan:

   y(n) : sinyal hasil pre-emphasis pada data ke-n

   s(n) : nilai sinyal pada data ke-n α :

   konstanta filter pre-emphasis

2. *Frame Blocking*

   Frame blocking adalah proses dimana sinyal audio dibagi menjadi beberapa potongan kecil (disebut dengan *frame*) yang saling tumpang tindih satu sama lain [9]. *Frame blocking* penting untuk dilakukan karena sinyal suara termasuk kedalam jenis sinyal yang terus berubah seiring berjalannya waktu. Sehingga dengan membagi sinyal kedalam *frame-frame* kecil diasumsikan bahwa sinyal tidak mengalami perubahan yang signifikan [9]. Persamaan *frame blocking* sebagai berikut:

   $$F = \left| N_f - \left((100\% - O) \times N_f \right) \right| N_s$$ ..... (2.2)

   Keterangan:

   F : jumlah *frame* dalam suatu sinyal audio

N_S : jumlah data dalam suatu sinyal
N_f : jumlah data dalam satu *frame* (amplitudo)
O : *overlapping* yang ditentukan (%)

3. *Windowing*

Windowing adalah sebuah proses pada pemrosesan sinyal yang bertujuan untuk meminimalkan diskontinuitas saat proses *frame blocking* berlangsung [9]. Diskontinuitas terletak pada bagian awal dan akhir tiap *frame*. *Windowing* bertindak sebagai filter yang meningkatkan sinyal di bagian tengah dan meratakan sinyal di bagian tepi pada seluruh *frame* [9]. Persamaan *windowing* sebagai berikut:

$y_{w(n)} = y_n \times w_n$ ..... (2.3)

Keterangan:
$y_w(n)$: sinyal hasil proses *windowing* pada indeks ke-n (amplitudo)
y_n : sinyal audio hasil pre-emphasis indeks ke-n (amplitudo)
$w(n)$: nilai *window* pada indeks ke-n

4. *Discrete Fourier Transform*

Discrete Fourier Transform (DFT) merupakan metode transformasi yang digunakan untuk mendekomposisi suatu sinyal kompleks menjadi sinyal sinus sederhana [10]. DFT menjadikan sinyal yang sebelumnya berada pada domain waktu berubah menjadi domain frekuensi. DFT memiliki kekurangan yaitu membutuhkan waktu komputasi yang terlalu lama dan tidak efisien, sehingga dikembangkan metode *Fast Fourier Transform* (FFT) untuk menutupi kekurangan tersebut [8]. Salah satu metode FFT yaitu *Short Time Fourier Transform* (STFT), STFT bekerja dengan baik pada sinyal yang telah dipotong menjadi *frame-frame* kecil. Persamaan STFT sebagai berikut:

$S_{k,m} = \sum_{n=1}^{N} y_{w(n),m} \times e^{-ikn}, 1 \leq m \leq F, 1 \leq k \leq \frac{N_f}{2}$.....(2.4)

Keterangan:
$S_{k,m}$: nilai *fourier transform* pada *frequency discrete* ke-k dan *frame* ke-m
$y_{w(n)}$: data sampel sinyal hasil *windowing* pada *frame* ke-m dan indeks ke-n
m : *frame* ke-m
k : variabel *frequency bin*

5. *Mel-Scaled Filterbank*

Mel-scaled filterbank adalah proses *filtering* yang mengubah sinyal audio dari satuan frekuensi menjadi satuan mel. Mel adalah satuan ukuran berdasarkan telinga manusia dalam menerima frekuensi suara, di mana telinga manusia tidak cukup sensitif untuk mendeteksi suara di bawah 1000 Hz [11]. Pemetaan antara frekuensi (Hz) dengan skala mel *linier* ketika di bawah 1000 Hz dan logaritmik ketika di atas 1000 Hz. Persamaan untuk mendapat nilai mel sebagai berikut:

$mel(f) = 2595\ log_{10}(1 + {}^{f}/_{700})$ ..... (2.5)

Keterangan:
f : *frequency*

6. *Discrete Cosine Transform*

   Discrete Coisine Transform (DCT) pada dasarnya memiliki kesamaan dengan *inverse* DFT, namun hasil yang didapat dari proses DCT lebih mendekati *Principal Component Analysis* (PCA) [8]. DCT bertujuan untuk mendapatkan fitur atau ciri *cepstral* dari *log-mel powerspectrum* [12]. Persamaan untuk mendapat nilai DCT sebagai berikut:

$$C_{k,m} = \sum_{i=1}^{K} \log_{10} Y[m] cos \left[\frac{\left(k - \frac{1}{2}\right)\pi}{K} \right], k = 1,2, \dots, K \quad \text{..... (2.6)}$$

   Keterangan:

   $C_{k,m}$: MFCC pada *cepstral coefficient* ke-k dan *frame* ke-m
   $Y[m]$: hasil dari proses *filterbank* pada *frame* ke-*m*
   k : nilai *coefficient*
   K : jumlah *coefficient* yang ditetapkan

7. *Cepstral Liftering*

   Ketika *cepstral coefficient* diekstrak dari sinyal suara, umumnya *cepstral coefficient* yang lebih kecil akan sensitif terhadap *spectral slope* dan *cepstral coefficient* yang lebih besar sensitif terhadap *noise*, sehingga untuk mengurangi masalah sensitifitas tersebut maka dilakukan *Cepstral Liftering* [13]. *Cepstral Liftering* meningkatkan akurasi untuk *pattern matching*, baik *speaker recognition* maupun *speech recognition* [6]. *Cepstral liftering* dapat diperoleh menggunakan sebagai berikut:

$$CL_{k,m} = \left(1 + \sin\left(\frac{\pi(c+1)}{C}\right) \times \frac{C}{2}\right) \times C_{k,m} \quad \text{..... (2.7)}$$

   Keterangan:

   $CL_{k,m}$: nilai *cepstral liftering* pada *cepstral coefficient* ke-*k* dan *frame* ke-*m*
   C : jumlah *cepstral coefficients*
   c : indeks *cepstral coefficient*

2.4. Statistik Dasar

Banyaknya data yang dihasilkan dari proses ekstraksi ciri bergantung pada jumlah *coefficient* MFCC serta durasi *file* audio. Perbedaan jumlah data akan mempengaruhi model *machine learning* dalam mempelajari pola, sehingga diperlukan suatu metode tambahan untuk menyamakan jumlah data yang ada. Salah satu metode tersebut adalah dengan menghitung statistik pada tiap *cepstral coefficient* yang dihasilkan. Statistik yang digunakan pada penelitian ini yaitu, *mean, max, median, min,* dan standar deviasi.

2.5. Support Vector Machine

Support Vector Machine (SVM) adalah metode *supervised learning* yang dikhususkan untuk menyelesaikan permasalahan dua kelas [13]. Namun terdapat beberapa pendekatan yang dapat digunakan untuk menyelesaikan permasalahan banyak kelas, seperti *one vs all* dan *one vs one*. SVM memiliki beberapa kernel yang berfungsi untuk perkalian dot di ruang dimensi tinggi seperti kernel *linear, rbf,* dan *polynomial.* SVM berusaha mencari *Optimal Separating Hyperplane* (OSH) sebuah garis yang memisahkan dataset secara *linear* dari berbagai kelas.

Persamaan untuk SVM secara umum sebagai berikut:

$$wx + b = 0 \quad \text{..... (2.8)}$$

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339

Keterangan:
w : vektor tegak lurus batas pemisah
x : data pada garis *margin*
b : *coefficient*

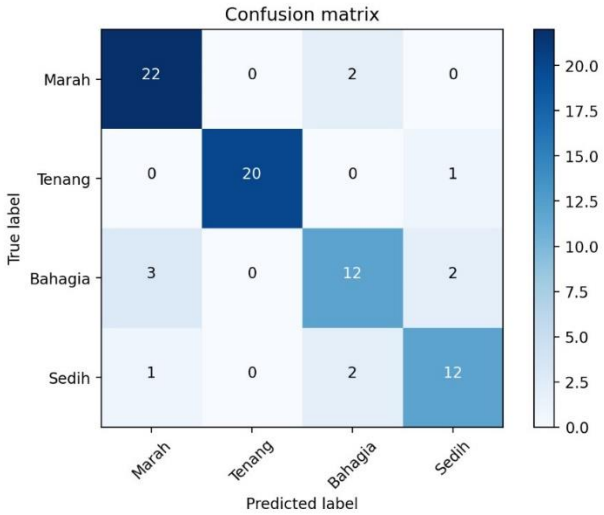
2.6. Pengujian Model

Pengujian model pada penelitian ini yaitu menggunakan *confusion matrix. Confusion matrix* banyak digunakan untuk mengetahui performa dari model *machine learning.* Performa meliputi *accuracy, precision,* dan *recall.*

3. Hasil dan Pembahasan

Jumlah data secara keseluruhan yang digunakan pada penelitian ini sebanyak 384 data untuk setiap gender, dengan pembagian data latih dan data uji sebesar 80% dan 20%. Sehingga total data latih dan uji berjumlah 307 data dan 77 data. Pengujian parameter *tuning* pada MFCC dan SVM menghasilkan akurasi yang berbeda antara laki-laki dan wanita. Akurasi terbaik yang dihasilkan pada gender laki-laki sebesar 85,71%. Sedangkan akurasi terbaik yang dihasilkan pada gender wanita sebesar 92,21%. Akurasi tersebut didapat dengan menghitung jumlah emosi yang diprediksi benar dibagi dengan jumlah keseluruhan data. Hasil tersebut kemudian dapat divisualisasikan dalam bentuk *confusion matrix.*

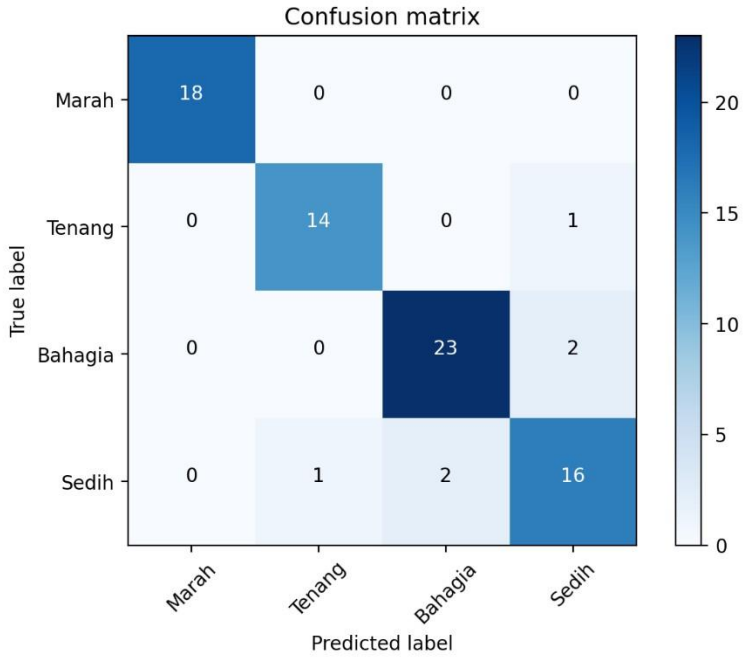
Confusion matrix pengujian pada laki-laki dengan menggunakan parameter hasil proses parameter *tuning* ditunjukkan seperti pada Gambar 3. Model *machine learning* SVM pada lakilaki mampu mengklasifikasikan emosi marah sebanyak 22 data dari 24 data, mengklasifikasikan emosi tenang sebanyak 20 data dari 21 data, mengklasifikasikan emosi bahagia sebanyak 12 data dari 17 data, dan mengklasifikasikan emosi sedih sebanyak 12 data dari 15 data. Akurasi terbaik pada laki-laki diperoleh dengan kombinasi parameter filter *pre-emphasize* sebesar 0.95, panjang *frame* sebesar 0.023 ms, *overlap frame* yang bersebelahan sebesar 40%, jumlah filter *mel* pada proses *mel-scaled filterbank* sebesar 24 mel, jumlah *cepstral coefficient* sebesar 24 *coefficient* dan nilai ‘C’ pada SVM sebesar 0,01.



Gambar 3. Confusion Matrix Pengujian Pria

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339

Confusion matrix pengujian pada wanita dengan menggunakan parameter hasil proses parameter *tuning* ditunjukkan seperti pada Gambar 4. Model *machine learning* SVM pada wanita mampu mengklasifikasikan emosi marah dengan sempurna sebanyak 18 data, mengklasifikasikan emosi tenang sebanyak 14 data dari 15 data, mengklasifikasikan emosi bahagia sebanyak 23 data dari 25 data, dan mengklasifikasikan emosi sedih sebanyak 16 data dari 19 data. Akurasi terbaik pada wanita diperoleh dengan kombinasi parameter filter *preemphasize* sebesar 0.95, panjang *frame* sebesar 0.023 ms, *overlap frame* yang bersebelahan sebesar 40%, jumlah filter *mel* pada proses *mel-scaled filterbank* sebesar 24 mel, dan jumlah *cepstral coefficient* sebesar 13 *coefficient* dan nilai 'C' pada SVM sebesar 0,01.



Gambar 4. Confusion Matrix Pengujian Wanit

Dari kedua hasil pengujian parameter *tuning* antara laki-laki dan wanita terdapat kesamaan nilai parameter pada semua parameter pengujian, kecuali pada jumlah *cepstral coefficient.* Jumlah *cepstral coefficient* pada laki-laki sebanyak 24 *coefficient* sedangkan jumlah *cepstral coefficient* pada wanita sebanyak 13 *coefficient.*

4. Kesimpulan dan Saran

Berdasarkan penelitian yang dilakukan, terdapat kesimpulan sebagai berikut, kombinasi metode MFCC dengan SVM mampu digunakan untuk klasifikasi emosi berdasarkan data masukan berupa intonasi suara dengan akurasi sebesar 85,71% pada laki-laki dan 92,21% pada wanita. Perbedaan akurasi yang didapat antara model laki-laki dan wanita disebabkan karena data yang digunakan berbeda. Model laki-laki dilatih dengan data suara laki-laki dan model wanita dilatih dengan data suara wanita, hal ini dilakukan karena laki-laki dan wanita memiliki rentang frekuensi suara yang berbeda. Kesimpulan ini didapat ketika digunakan pada dataset penelitian RAVDESS.

Selain itu penelitian ini jauh dari kata sempurna dan masih banyak kekurangan di dalamnya sehingga perlu adanya perbaikan untuk menghasilkan penelitian yang lebih baik di masa yang akan datang. Adapun saran yang dapat diberikan untuk penelitian selanjutnya yaitu:

1. Menggabungkan beberapa data penelitian lain dengan bahasa yang sama untuk memperbanyak data penelitian sehingga model yang dilatih lebih baik.
2. Menggunakan metode lain yang dapat membedakan ciri antara pria dan wanita lebih spesifik.
3. Mennggabungkan beberapa metode ekstraksi ciri seperti MFCC dengan LPC dimana MFCC dikembangkan berdasarkan sistem pendengaran manusia sedangkan LPC dikembangkan berdasarkan filter produksi saluran vocal manusia.

Daftar Pustaka

[1] Helmiyah, S., Riadi, I., Umar, R., & Hanif, A. (2021). Speech Classification to Recognize Emotion Using Artificial Neural Network. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, *7*(1), 12–17. https://doi.org/10.23917/khif.v7i1.11913

[2] Ancilin, J., & Milton, A. (2021). Improved speech emotion recognition with Mel frequency magnitude coefficient. *Applied Acoustics*, *179*, 108046. https://doi.org/10.1016/j.apacoust.2021.108046

[3] Muljono, Prasetya, M. R., Harjoko, A., & Supriyanto, C. (2019). Speech Emotion Recognition of Indonesian Movie Audio Tracks based on MFCC and SVM. *Proceedings of the 4th International Conference on Contemporary Computing and Informatics, IC3I 2019*, 22–25. https://doi.org/10.1109/IC3I46837.2019.9055509

[4] Aini, Y. K., Santoso, T. B., & Dutono, D. T. (2021). Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia. Jurnal Komputer Terapan, 7(1), 143–152. https://jurnal.pcr.ac.id/index.php/jkt/

[5] Al Dujaili, M. J., Ebrahimi-Moghadam, A., & Fatlawi, A. (2021). Speech emotion recognition based on SVM and KNN classifications fusion. *International Journal of Electrical and Computer Engineering*, *11*(2), 1259–1264. https://doi.org/10.11591/ijece.v11i2.pp1259-1264

[6] Putra, D. dan Adi, R., 2011. Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW. Biometrika, Universitas Udayana, 2(1), hal.8–21.

[7] Arifin, C., & Junaedi, H. (2018). Emotion Sound Classification with Support Vector Machine Algorithm. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 3(2), 181–190. https://doi.org/10.22219/kinetik.v3i2.610

[8] Heriyanto, H., Hartati, S., & Putra, A. E. (2018). Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-Qur'an. Telematika, 15(2), 99. https://doi.org/10.31315/telematika.v15i2.3123

[9] Oday, K. (2018). Frame Blocking and Windowing Speech Signal. *Journal of Information, Communication, and Intelligence Systems*, *4*(No. 5).

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v21i3.12339

https://www.researchgate.net/publication/331635757

[10] Ashshiddieqy, M. H., Jondri, & Rizal, A. (2020). Klasifikasi Suara Paru Dengan Convolutional Neural Network (CNN). *EProceedings of Engineering*, *7*(2), 8506–8512.

[11] Altayeb, M., & Al-Ghraibah, A. (2022). Classification of three pathological voices based on specific features groups using support vector machine. *International Journal of Electrical and Computer Engineering*, *12*(1), 946–956. https://doi.org/10.11591/ijece.v12i1.pp946-956

[12] Ramashini, M., Abas, P. E., Mohanchandra, K., & de Silva, L. C. (2022). Robust cepstral feature for bird sound classification. *International Journal of Electrical and Computer Engineering*, *12*(2), 1477–1487. https://doi.org/10.11591/ijece.v12i2.pp1477-1487

[13] Paylakhi, S. H., & Shirazi, J. (2015). Phone Recognition on TIMIT Database Using MelFrequency Cepstral Coefficients and Support Vector Machine Classifier. In *IJISETInternational Journal of Innovative Science*, *Engineering & Technology* (Vol. 2, Issue 1). www.ijiset.com