



Emotion Recognition Based on Acoustic Features of Speech Using Support Vector Machine

Pengenalan Emosi Berdasarkan Fitur Akustik Suara Menggunakan Metode Support Vector Machine

Veronika Lia Listiyaningsih^{1*}, Susi Marianingsih²

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Manajemen,
Universitas Sains dan Teknologi Jayapura, Indonesia

E-Mail: ¹veronikalistiyaningsih@gmail.com, ²ssmarianingsih@gmail.com

Received Apr 11th 2026; Revised Jun 04th 2026; Accepted Jun 18th 2026; Available Online Jun 25th 2026

Corresponding Author: Veronika Lia Listiyaningsih

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Speech emotion recognition is an important field in speech signal processing and artificial intelligence that aims to identify a person's emotional state based on acoustic speech characteristics. This study develops a speech emotion classification model using the Support Vector Machine (SVM) algorithm with a combination of Mel-Frequency Cepstral Coefficients (MFCC), pitch, and energy features to represent spectral characteristics, fundamental frequency, and speech intensity. The RAVDESS dataset, consisting of 1,440 audio samples across eight emotion classes, was used in this study. After the data cleaning process, 1,245 audio samples were retained for analysis. The research workflow included audio preprocessing, feature extraction and normalization, training-testing data splitting, model training, hyperparameter tuning using GridSearchCV, and performance evaluation. Experimental results showed that the SVM model with the Radial Basis Function (RBF) kernel achieved the best accuracy of 82%, with macro-average precision, recall, and F1-score of 0.82. Model performance improved progressively from 40% in the initial model to 60% after data cleaning, 77% after feature extraction and normalization, and 82% following hyperparameter tuning. The combination of MFCC, pitch, and energy features outperformed individual features and partial feature combinations in speech emotion classification.

Keywords: *Acoustic Features, Emotion Detection, MFCC, Speech Emotion Recognition, Support Vector Machine*

Abstrak

Pengenalan emosi berbasis suara merupakan bidang penting dalam pengolahan sinyal suara dan kecerdasan buatan yang bertujuan untuk mengidentifikasi kondisi emosional seseorang berdasarkan karakteristik akustik. Penelitian ini mengembangkan model klasifikasi emosi menggunakan Support Vector Machine (SVM) dengan kombinasi fitur Mel-Frequency Cepstral Coefficients (MFCC), pitch, dan energy untuk merepresentasikan karakteristik spektral, frekuensi dasar, dan intensitas suara. Dataset yang digunakan adalah RAVDESS yang terdiri atas 1.440 data audio dengan delapan kelas emosi. Setelah proses data cleaning, sebanyak 1.245 data digunakan dalam penelitian. Tahapan meliputi preprocessing audio, ekstraksi dan normalisasi fitur, pembagian data latih dan uji, pelatihan model, hyperparameter tuning menggunakan GridSearchCV, serta evaluasi performa. Hasil pengujian menunjukkan bahwa SVM dengan kernel Radial Basis Function (RBF) mencapai akurasi terbaik sebesar 82%, dengan nilai macro-average precision, recall, dan F1-score masing-masing sebesar 0,82. Performa model meningkat dari 40% pada model awal menjadi 60% setelah data cleaning, 77% setelah ekstraksi dan normalisasi fitur, serta 82% setelah hyperparameter tuning. Kombinasi MFCC, pitch, dan energy memberikan performa terbaik dibandingkan dengan penggunaan fitur tunggal maupun kombinasi parsial.

Kata kunci: *Fitur Akustik, MFCC, Pengenalan Emosi, Pengenalan Emosi Suara, Support Vector Machine*

1. PENDAHULUAN

Emosi menjadi permasalahan terpenting dalam berkomunikasi karena dapat mencerminkan kondisi psikologis, respons afektif, dan keadaan mental seseorang [1], [2]. Suara manusia tidak hanya membawa informasi linguistik, tetapi juga memuat informasi paralinguistik seperti intonasi, tekanan suara, ritme, frekuensi dasar, dan intensitas suara yang berkaitan dengan ekspresi emosi [2], [3]. Informasi tersebut dapat dimanfaatkan dalam *speech emotion recognition* atau pengenalan emosi berbasis suara untuk



mengidentifikasi kategori emosi seseorang secara otomatis [1], [3]. Oleh karena itu, pengenalan emosi berbasis suara menjadi salah satu bidang penting dalam pengolahan sinyal dan kecerdasan buatan, terutama untuk mendukung sistem interaksi manusia dan komputer, layanan kesehatan digital, pembelajaran adaptif, serta sistem analisis emosi berbasis komputasi [4].

Pengenalan emosi secara konvensional umumnya dilakukan melalui observasi langsung, wawancara, kuesioner, atau penilaian subjektif oleh manusia [2], [5]. Pendekatan tersebut memiliki keterbatasan karena bergantung pada persepsi penilai, membutuhkan waktu yang relatif lama, dan sulit diterapkan pada sistem otomatis berskala besar. Perkembangan *machine learning* memungkinkan proses pengenalan emosi dilakukan secara lebih objektif melalui pembelajaran pola dari data suara [2], [6]. Melalui pendekatan berbasis data, model klasifikasi dapat mempelajari hubungan antara karakteristik akustik suara dan kategori emosi tertentu secara otomatis [2].

Dalam pengenalan emosi berbasis suara, ekstraksi fitur menjadi tahap penting karena kualitas fitur sangat menentukan kemampuan model dalam membedakan kelas emosi [2], [3]. Diantara fitur yang banyak dipakai adalah *Mel-Frequency Cepstral Coefficients* atau MFCC karena mampu merepresentasikan karakteristik spektral suara berdasarkan pendekatan persepsi pendengaran manusia [3], [7]. Selain MFCC, fitur *pitch* juga penting karena menggambarkan frekuensi dasar suara yang berkaitan dengan tinggi rendahnya nada dan perubahan intonasi [2], [8]. Fitur *energy* digunakan untuk merepresentasikan kuat atau lemahnya intensitas suara yang sering berubah sesuai kondisi emosional pembicara [12]. Kombinasi MFCC, *pitch*, dan *energy* dapat memberikan representasi akustik yang lebih lengkap karena mencakup aspek spektral, frekuensi dasar, dan intensitas sinyal suara [3].

Berbagai penelitian menunjukkan bahwa pemilihan fitur dan metode klasifikasi sangat memengaruhi keberhasilan sistem pengenalan emosi berbasis suara [1]. Ancilin dan Milton menunjukkan bahwa pengembangan fitur berbasis *Mel* dapat meningkatkan kualitas representasi sinyal suara dalam pengenalan emosi [2]. *Speech emotion recognition* dapat dikembangkan menggunakan metode *machine learning* konvensional maupun *deep learning*, dengan performa yang dipengaruhi oleh kualitas data, fitur, dan strategi klasifikasi [3]. Meskipun pendekatan *deep learning* banyak digunakan, metode *machine learning* konvensional tetap relevan karena lebih sederhana, efisien, dan dapat diterapkan pada dataset berukuran terbatas [5].

Beberapa algoritma *machine learning* yang umum digunakan dalam klasifikasi emosi suara adalah K-Nearest Neighbors atau KNN, Random Forest, dan SVM [6], [7]. KNN mengklasifikasikan data berdasarkan kedekatan jarak antar sampel, tetapi performanya dapat menurun ketika data memiliki dimensi fitur tinggi atau distribusi kelas yang saling tumpang tindih. SVM bekerja dengan membentuk *hyperplane* optimal untuk memecahkan data pada kelas tertentu dengan margin terbaik. Pada data yang memiliki pola non-linear, SVM dapat dikombinasikan dengan kernel *Radial Basis Function* atau RBF untuk meningkatkan kemampuan pemisahan antar kelas [9].

Dataset RAVDESS banyak digunakan sebagai dataset standar dalam penelitian pengenalan emosi berbasis suara karena menyediakan data audio dengan ekspresi emosi yang terstruktur dan melibatkan aktor profesional [10]. Dataset ini memuat delapan kategori emosi, yaitu *neutral*, *calm*, *happy*, *sad*, *angry*, *fearful*, *disgust*, dan *surprised*. Meskipun RAVDESS direkam dalam kondisi relatif terkontrol, proses *preprocessing* dan *data cleaning* tetap diperlukan untuk menjaga kualitas data yang digunakan dalam pelatihan model. Perbedaan durasi sinyal, variasi intensitas, bagian hening, dan kemungkinan data yang tidak optimal dapat memengaruhi proses ekstraksi fitur dan performa klasifikasi.

Tantangan utama dalam pengenalan emosi berbasis suara adalah kemiripan karakteristik akustik antar emosi tertentu. Emosi seperti *neutral* dan *calm* cenderung memiliki intensitas serta variasi *pitch* yang rendah sehingga sulit dibedakan oleh model. Emosi seperti *fearful* dan *surprised* juga dapat memiliki kemiripan dalam perubahan *pitch* dan *energy* yang tinggi sehingga berpotensi menimbulkan kesalahan klasifikasi. Selain itu, performa model dipengaruhi oleh pemilihan fitur, skala data, parameter model, dan strategi. Berdasarkan kajian penelitian terdahulu, sebagian besar penelitian *speech emotion recognition* berfokus pada penggunaan fitur MFCC secara tunggal atau menerapkan pendekatan *deep learning* yang membutuhkan sumber daya komputasi relatif tinggi. Selain itu, penelitian yang secara khusus mengevaluasi kontribusi kombinasi fitur MFCC, *pitch*, dan *energy* menggunakan SVM dengan kernel RBF pada dataset RAVDESS masih relatif terbatas. Sebagian besar penelitian juga lebih menitikberatkan pada hasil akurasi akhir tanpa menganalisis secara rinci pengaruh setiap tahapan *preprocessing*, normalisasi fitur, dan *hyperparameter tuning* terhadap peningkatan performa model. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan mengevaluasi efektivitas kombinasi fitur akustik serta optimasi model SVM-RBF secara sistematis dalam proses pengenalan emosi berbasis suara.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada pengenalan emosi berbasis suara menggunakan kombinasi fitur MFCC, *pitch* dan *energy* pada dataset RAVDESS. Metode utama yang digunakan adalah SVM dengan kernel RBF karena mampu menangani pola non-linear pada fitur akustik suara [11]. Untuk mengetahui efektivitas metode yang digunakan, penelitian ini membandingkan performa SVM RBF dengan KNN, Random Forest, dan SVM Linear. Selain itu, penelitian ini juga menganalisis

pengaruh kombinasi fitur MFCC, MFCC + *pitch*, MFCC + *energy*, dan MFCC + *pitch* + *energy* terhadap performa klasifikasi. Penelitian ini berharap dapat menjadikan model pengenalan emosi berbasis suara yang sederhana dan efisien melalui pemilihan fitur akustik dan metode klasifikasi yang tepat, sekaligus memberikan gambaran mengenai pengaruh *preprocessing*, kombinasi fitur, dan *tuning hyperparameter* terhadap performa klasifikasi.

2. DATASET DAN METODE

2.1 Dataset

Pada penelitian ini, dataset yang dipakai adalah RAVDESS yang terdiri dari 1440 data audio dengan delapan kelas emosi [10]. Namun, tidak seluruh data dapat digunakan secara langsung karena terdapat beberapa file audio yang tidak memenuhi kriteria kualitas yang dibutuhkan dalam proses ekstraksi fitur. Berikut dirangkum dalam tabel kategori emosi yang ditunjukkan pada Tabel 1.

Tabel 1. Kategori Emosi

Jenis Emosi	Jumlah Data
Neutral	96
Calm	192
Happy	192
Sad	192
Angry	192
Fearful	192
Disgust	192
Surprised	192
TOTAL	1.440

Dataset yang digunakan direkam oleh 24 aktor profesional dengan dua tingkat intensitas emosi, yaitu normal dan kuat, kecuali emosi *neutral* yang hanya memiliki satu tingkat intensitas sehingga jumlah datanya relatif lebih sedikit. Meskipun demikian, distribusi data secara umum cukup seimbang dalam proses pembelajaran mesin [9].

Sebelum digunakan, dilakukan proses *data cleaning* untuk menghapus data yang tidak memenuhi kriteria, seperti durasi terlalu pendek, noise berlebih, serta data yang gagal diproses saat ekstraksi fitur. Setelah proses tersebut, jumlah data yang digunakan menjadi 1245 sampel. Tahapan ini bertujuan untuk menguatkan kualitas data hingga cara belajar bisa lebih optimal. Setelah proses cleaning, jumlah data berkurang dari 1440 menjadi 1245 data. Hasil proses cleaning dijelaskan pada Tabel 2.

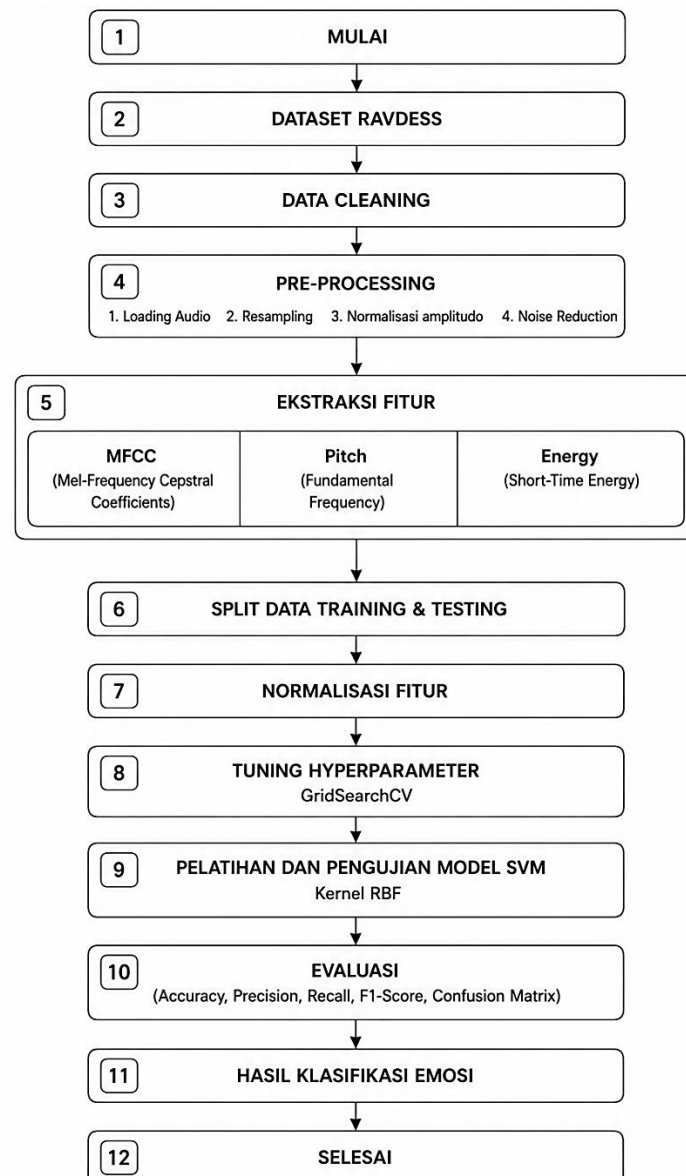
Tabel 2. Hasil proses *data cleaning*

Kelas Emosi	Data Awal	Data Dihapus	Data Akhir	Alasan Penghapusan
Neutral	96	10	86	Durasi pendek
Calm	192	25	167	Noise, durasi pendek
Happy	192	24	168	Noise, durasi pendek
Sad	192	23	169	Fitur tidak Valid
Angry	192	26	166	Noise tinggi
Fearful	192	24	168	Noise tinggi
Disgust	192	31	161	Noise, anomali fitur
Surprised	192	32	160	Durasi pendek
Total	1440	195	1245	

Distribusi data setelah proses *data cleaning* masih relatif seimbang antar kelas, meskipun terdapat sedikit pembeda hasil lainnya. Perbedaan ini tidak signifikan dan masih dalam batasan, sehingga tidak memberikan bias yang berarti terhadap proses pelatihan model. Perlu diperhatikan bahwa nilai *support* pada classification report menunjukkan jumlah data pada data uji (testing set), bukan seluruh dataset. Pada penelitian ini, dataset dibagi dengan rasio 80:20, sehingga jumlah data uji yang dipakai adalah sekitar 249 data dari total 1245 data.

2.2 Metode

Tahapan penelitian ini dirancang secara terstruktur untuk menggambarkan alur proses dalam pengolahan data suara hingga proses klasifikasi emosi. Dimulai dari tahap segmentasi dan *preprocessing* untuk meningkatkan kualitas sinyal audio, kemudian dilanjutkan dengan ekstraksi fitur akustik yang meliputi MFCC, *pitch*, dan *energy*. Setelah itu, dilakukan pembagian data serta proses klasifikasi menggunakan metode Support Vector Machine (SVM) dengan kernel RBF. Setiap tahapan saling berkaitan dalam membentuk sistem pengenalan emosi yang optimal. Secara keseluruhan, tahapan penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Proses Klasifikasi

2.2.1. Segmentasi Audio

Segmentasi audio dilakukan dengan membagi sinyal suara menjadi *frame* pendek untuk menangani sifat non-stasioner sinyal. Pendekatan ini memungkinkan analisis lokal yang lebih akurat dan stabil, dengan panjang *frame* umumnya 20–30 ms serta menggunakan *overlap* untuk menjaga kontinuitas sinyal [2].

2.2.2. Preprocessing

Tahap *preprocessing* dipakai untuk mempersiapkan data audio sebelum proses ekstraksi fitur. Proses ini bertujuan untuk memastikan kualitas sinyal audio tetap konsisten dan dapat diolah secara optimal oleh model. Pada penelitian ini, *preprocessing* dilakukan menggunakan beberapa *library* Python, yaitu *librosa* untuk pengolahan audio, *numpy* untuk operasi numerik, serta *scikit-learn* untuk normalisasi data.

Adapun tahapan *preprocessing* yang dilakukan dalam penelitian ini meliputi:

1. *Loading audio*: merupakan proses membaca file audio ke dalam sistem dan mengubahnya menjadi sinyal digital yang berkelanjutan. [10].
2. *Resampling*: *Resampling* dilakukan untuk menyeragamkan frekuensi sampling seluruh sinyal audio menjadi 22050 Hz, sehingga setiap data memiliki resolusi temporal yang sama dan dapat dibandingkan secara konsisten. Proses ini melibatkan interpolasi dan filtering untuk mencegah aliasing [11].

Secara matematis, proses resampling dapat dinyatakan sebagai:

$$x_r[n] = x\left(n \cdot \frac{f_s}{f_{new}}\right) \quad (1)$$

Keterangan:

$x[n]$: sinyal asli
 $x_r[n]$: sinyal hasil resampling
 f_s : frekuensi sampling awal
 f_{new} : frekuensi sampling baru

Proses ini dilakukan menggunakan parameter $sr=22050$ pada fungsi *librosa.load()*.

3. *Normalisasi amplitudo*: Normalisasi amplitudo bertujuan untuk menyamakan skala amplitudo sinyal agar berada dalam rentang tertentu, sehingga mengurangi perbedaan intensitas antarsinyal dan meningkatkan stabilitas proses ekstraksi fitur. Rumus normalisasi amplitudo dinyatakan pada persamaan 2.

$$x_{norm}[n] = \frac{x[n]}{\max(|x[n]|)} \quad (2)$$

Keterangan:

$x[n]$: sinyal asli
 $x_{norm}[n]$: sinyal hasil normalisasi

4. *Noise reduction*: *Noise reduction* merupakan proses untuk mengurangi gangguan atau derau (*noise*) pada sinyal audio sehingga sinyal utama dapat dianalisis dengan lebih akurat. Dalam penelitian ini digunakan metode spectral subtraction, yang bekerja dengan mengurangi spektrum noise dari spektrum sinyal. Secara matematis, metode ini dirumuskan pada persamaan 3.

$$|Y(\omega)| = |X(\omega)| - |N(\omega)| \quad (3)$$

Keterangan:

$Y(\omega)$: spektrum sinyal bersih
 $X(\omega)$: spektrum sinyal asli
 $N(\omega)$: spektrum noise

5. *Trimming silence*: *Trimming silence* dilakukan untuk menghilangkan bagian hening pada awal dan akhir sinyal audio yang tidak mengandung informasi penting. Proses ini biasanya dilakukan menggunakan threshold energy. Secara matematis, proses ini dapat dinyatakan pada persamaan 4.

$$x_{trim}[n] = x[n] \cdot u(E[n] - T) \quad (4)$$

Keterangan:

$x[n]$: sinyal asli
 $x_{trim}[n]$: sinyal setelah *trimming*
 $E[n]$: sinyal energi
 T : *threshold*
 $u(\cdot)$: fungsi step (*Heaviside*)

6. *Framing dan Windowing*: Proses *framing* dilakukan untuk membagi sinyal menjadi segmen-segmen kecil agar dapat dianalisis dalam domain waktu dan frekuensi.

Parameter yang digunakan pada setiap tahapan *preprocessing* dan pemodelan dirangkum dan ditunjukkan pada Tabel 3.

Tabel 3. Detail Parameter

Tahap	Parameter yang Digunakan
Resampling	<i>Sampling rate</i> = 22050 Hz
Trimming	<i>Metode librosa.effects.trim()</i> dan <i>Threshold</i> : <i>top_db</i> =20
Framing	<i>frame length</i> (<i>n_fft</i>)=2048, <i>Hop length</i> =512, <i>Overlap</i> =75%
MFCC	<i>N_mfcc</i> =13, <i>n_fft</i> =2048, <i>hop_length</i> =512, <i>n_mels</i> =40
Split data	Rasio 80:20 (<i>train:test</i>) menggunakan <i>Stratified Split</i>
Tuning	$C = [0.1, 1, 10, 100]$, <i>gamma</i> =[scale, auto, 0.001, 0.01, 0.1, 1], Kernel=RBF
Validasi	<i>hold-out validation</i>

2.3 Ekstraksi Fitur

Ekstraksi fitur merupakan tahapan penting dalam pengolahan sinyal suara yang bertujuan untuk mengubah sinyal audio menjadi representasi numerik yang lebih ringkas dan informatif. Fitur yang dihasilkan harus mampu merepresentasikan karakteristik penting dari sinyal suara sehingga dapat digunakan secara optimal dalam proses klasifikasi. Pada penelitian ini, fitur yang digunakan terdiri dari *Mel-Frequency Cepstral Coefficients (MFCC)*, *pitch*, dan *energy* [2], [3], [12].

2.3.1. Mel-Frequency Cepstral Coefficients (MFCC)

MFCC menjadi metode ekstraksi fitur umum yang digunakan dalam pengolahan sinyal suara karena mampu memaparkan karakteristik spektral berdasarkan tanggapan pendengaran manusia. MFCC bekerja dengan mengubah sinyal suara dari domain waktu ke domain frekuensi, kemudian memetakannya ke skala *Mel* yang menyerupai cara manusia mendengar suara. MFCC efektif dalam menangkap informasi penting dari sinyal suara sehingga banyak digunakan dalam pengenalan suara dan emosi [4]. Nilai MFCC yang digunakan merupakan nilai rata-rata (mean) dari seluruh frame untuk setiap koefisien, sehingga menghasilkan representasi yang lebih ringkas.

Proses MFCC melibatkan beberapa tahapan, di antaranya transformasi Fourier yang ditunjukkan pada persamaan 5.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N} \quad (5)$$

Kemudian dilakukan pemetaan ke skala *Mel*, yang ditunjukkan pada persamaan 6.

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (6)$$

Dilanjutkan dengan Discrete Cosine Transform (DCT), yang ditunjukkan pada persamaan 7.

$$C_n = \sum_{k=1}^K \log(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (7)$$

2.3.2. Pitch

Pitch merupakan frekuensi dasar dari sinyal suara, yang berperan penting dalam merepresentasikan intonasi dan emosi, karena perubahan emosi sering kali diikuti oleh perubahan frekuensi suara. *Pitch* merupakan salah satu indikator utama dalam analisis emosi berbasis suara karena berkaitan langsung dengan ekspresi vokal manusia [12]. Perhitungan frekuensi dasar dinyatakan pada persamaan 8.

$$f_0 = \frac{1}{T} \quad (8)$$

Keterangan :

f_0 : Frekuensi dasar
 T : Periode sinyal

2.3.3. Energy

Energy merupakan fitur yang menunjukkan tingkat intensitas atau kekuatan sinyal suara. *Energy* berhubungan dengan keras atau lemahnya suara yang dihasilkan, yang juga mencerminkan kondisi emosional seseorang. *energy* merupakan fitur penting dalam membedakan emosi seperti marah, senang, dan sedih [12]. *Energy* dihitung sebagai jumlah kuadrat amplitudo sinyal, yang ditunjukkan pada persamaan 9.

$$E = \sum_{n=1}^N |x(n)|^2 \quad (9)$$

Keterangan:

$x[n]$: Sinyal audio
 N : Jumlah sampel

Penggunaan ketiga fitur tersebut secara bersamaan diharapkan mampu memberikan representasi yang lebih komprehensif terhadap karakteristik suara, sehingga dapat meningkatkan performa model dalam mengklasifikasikan emosi. Ringkasan dan jumlah dimensi fitur yang digunakan dalam penelitian ini dijelaskan pada Tabel 4.

Tabel 4. Penggunaan Fitur

Fitur	Jumlah Dimensi
MFCC	13

Fitur	Jumlah Dimensi
Pitch/F0	1
Energy/RMS	1
Total	15

2.4. Split Data

Dataset dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*). Pembagian ini bertujuan untuk melatih dan menguji performa model terhadap data yang tak pernah dilatih sebelumnya.

2.5. Support Vector Machine (SVM)

Support Vector Machine adalah algoritma *machine learning* yang berupaya menemukan *hyperplane* optimal sedemikian rupa hingga jarak antara titik data terdekat dari setiap kelas dan *hyperplane* tersebut menjadi maksimum [11]. Dalam kasus data yang tidak dapat dipisahkan secara linier (*non-linearly separable data*), vektor input dari SVM dapat dipetakan ke ruang berdimensi lebih tinggi, sehingga data tersebut dapat dipisahkan secara linier.

Namun, pemetaan vektor input ke ruang berdimensi lebih tinggi memiliki kompleksitas komputasi yang sangat tinggi. Untuk mengatasi kompleksitas ini, digunakan fungsi kernel. Fungsi kernel ini mengurangi kompleksitas dimensi hingga kembali ke dimensi pada ruang input. Tiga jenis kernel yang umum digunakan [9] ditampilkan pada Tabel 5.

Tabel 5. Jenis Kernel

Kernel	Representasi Matematis
<i>Polynomial kernel</i> dengan derajat p (<i>Polynomial</i>)	$K(x_i, x) = (x_i * x + 1)$
<i>Gaussian radial basis function kernel</i> (<i>RBF</i>)	$K(x_i, x) = e^{-\frac{\ x_i - x\ ^2}{2\sigma^2}}$
<i>Sigmoid kernel</i> (<i>Sigmoid</i>)	$K(x_i, x) = \tanh(K x_i * x - \delta)$

2.6. Tuning hyperparameter

Tuning hyperparameter merupakan proses optimasi *hyperparameter* yang tidak dipelajari secara langsung dari data, melainkan ditentukan sebelum proses pelatihan [9]. Pada penelitian ini, tuning dilakukan pada parameter C dan γ pada model SVM dengan kernel RBF. Parameter C berfungsi untuk mengatur keseimbangan antara margin maksimum dan kesalahan klasifikasi, sedangkan parameter γ mengontrol tingkat kompleksitas model dalam memisahkan data. Proses tuning dilakukan dengan mencoba beberapa kombinasi nilai parameter secara sistematis untuk memperoleh hasil klasifikasi terbaik. Berikut adalah parameter yang diuji dan hasil terbaik yang diperoleh ditunjukkan pada Tabel 6.

Tabel 6. Hasil Tuning

Parameter	Nilai yang Diuji	Nilai Terbaik
C	0.1, 1, 10, 100	10
γ	scale, auto, 0.001, 0.01, 0.1, 1	0.01
Kernel	RBF	
Validation	5-fold Cross Validation (GridSearchCV)	

Parameter C berfungsi untuk mengontrol *trade-off* antara margin yang besar dan kesalahan klasifikasi. Nilai C yang lebih besar akan menghasilkan margin yang lebih kecil, namun dengan tingkat kesalahan yang lebih rendah. Parameter γ menentukan seberapa jauh pengaruh satu data terhadap data lainnya. Nilai γ yang kecil menghasilkan model yang lebih halus, sedangkan nilai yang besar dapat menyebabkan *overfitting*.

Berdasarkan hasil tuning, kombinasi parameter terbaik diperoleh pada $C=10$ dan $\gamma = 0.01$, yang menghasilkan performa model terbaik. Model SVM dengan parameter terbaik tersebut mampu mencapai accuracy sebesar 82% pada data uji.

2.7. Evaluasi Model

Evaluasi utama dilaksanakan menggunakan train-test split dengan rasio 80:20 secara stratified. Proses *tuning hyperparameter* dilakukan menggunakan *GridSearchCV* dengan *5-fold cross-validation* pada data pelatihan. Selain itu, untuk menguji stabilitas performa model terhadap variasi pembagian data, dilakukan *repeated hold-out* sebanyak 30 kali dengan nilai random state yang berbeda.

Model dievaluasi menggunakan beberapa metrik untuk mengukur kinerja klasifikasi dalam mengenali emosi berdasarkan suara. Evaluasi dilakukan terhadap hasil klasifikasi emosi menggunakan metode Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Metrik yang digunakan meliputi Confusion Matrix, Accuracy, Precision, Recall, dan F1-score [2], [6], [7].

2.8. Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi dan data aktual. Matriks ini terdiri dari empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) .

1. Accuracy

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan seluruh data dengan membandingkan jumlah prediksi dengan total data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

2. Precision

Precision menunjukkan tingkat ketepatan model dalam memprediksi data yang termasuk ke dalam kelas tertentu, yaitu perbandingan antara jumlah prediksi yang benar pada kelas positif dengan jumlah seluruh data yang diprediksi sebagai positif.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

3. Recall

Recall dipakai untuk pengukuran keahlian dalam mendeteksi keseluruhan data yang benar dari suatu kelas, dengan membandingkan jumlah prediksi benar pada kelas positif dengan seluruh data aktual pada kelas tersebut [15].

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

4. F1-Score

F1-score menjadi rata-rata harmonis diantara *precision* dan *recall* yang dipakai dalam pemberian pemaparan penyeimbangan performa model [15].

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (13)$$

Pada penelitian ini, evaluasi dilakukan pada kasus multi-class, sehingga perhitungan *precision*, *recall*, dan *F1-score* memakai pendekatan rata-rata seperti *macro average* atau *weighted average*.

3. HASIL DAN PEMBAHASAN

Proses ekstraksi fitur menghasilkan 15 fitur yang terdiri 13 fitur MFCC, 1 fitur *pitch*, dan 1 fitur *energy*. Contoh dataset fitur audio ditunjukkan pada Gambar 2.

	MFCC_1	MFCC_2	MFCC_3	MFCC_4	MFCC_5	MFCC_6	MFCC_7	MFCC_8	MFCC_9	MFCC_10	MFCC_11	MFCC_12	MFCC_13	Pitch	Energy	Emotion
0	-332.713318	20.409544	-24.743662	-12.518736	-15.651567	-15.802431	-6.951442	-10.635042	-7.456793	-5.074695	-3.823032	-4.884599	-3.884145	167.274146	0.062067	05
1	-317.358978	23.297577	-39.682808	-12.729723	-22.555910	-15.028669	-10.055988	-21.065582	-5.627684	2.649390	-3.437780	7.237745	1.696630	164.679756	0.050029	06
2	-697.178284	34.360626	-10.023999	2.262170	-5.140885	-3.702223	-10.113008	-13.919273	-5.329112	-3.939005	-5.058762	-4.667789	-3.529691	190.356059	0.002494	08
3	-797.798645	32.446926	-6.713991	6.684238	-4.769311	-5.839230	-9.006455	-10.979172	-6.628823	-4.447314	-5.281420	-9.067540	-4.236372	208.022021	0.000739	02
4	-795.281860	44.769058	-1.890999	13.641954	-0.082623	5.293716	-6.862940	-14.351247	-8.375112	-0.114325	-3.037451	-5.556241	-2.840406	146.497563	0.000581	04

Gambar 2. Contoh Dataset Fitur Audio

Dataset di atas menjadi data masukan bagi model SVM. Untuk mendapatkan hasil terbaik maka dilakukan pengujian kontribusi fitur terhadap *accuracy* klasifikasi dan pengujian untuk mengetahui pengaruh *preprocessing*, normalisasi fitur dan *tuning* pada *accuracy* klasifikasi.

3.1 Analisis Kontribusi Fitur Terhadap Accuracy Klasifikasi

Pengujian berikut untuk mengetahui pengaruh tiap fitur terhadap kinerja model. Beberapa kombinasi fitur yang berbeda menjadi input bagi klasifikasi SVM. Hasil pengujian *accuracy* untuk tiap kombinasi fitur ditunjukkan pada Tabel 7.

Tabel 7. Perbandingan accuracy untuk tiap kombinasi fitur

Fitur	Accuracy	Precision	Recall	F1-Score
MFCC	0.49	0.47	0.47	0.46
MFCC+Pitch	0.54	0.53	0.53	0.52
MFCC+ Energy	0.50	0.49	0.49	0.48
MFCC+Pitch+Energy	0.82	0.82	0.82	0.82

Hasil pengujian menunjukkan bahwa kombinasi fitur MFCC + pitch + energy memberikan performa terbaik dibandingkan dengan penggunaan fitur tunggal. MFCC menjadi fitur utama, sementara *pitch* dan *energy* berperan sebagai fitur pendukung untuk meningkatkan *accuracy*.

3.2 Pengaruh Preprocessing, Normalisasi fitur dan Tuning Pada Accuracy Klasifikasi

Hasil penelitian menunjukkan bahwa metode SVM dengan kernel RBF mampu mengklasifikasikan emosi suara dengan performa yang meningkat secara bertahap setelah dilakukan serangkaian tahapan *preprocessing* dan optimasi model. Ringkasan tahapan eksperimen dan peningkatan *accuracy* model disajikan pada Tabel 8.

Pada tahap awal, model menghasilkan *accuracy* sebesar 40%, yang menunjukkan bahwa performa klasifikasi masih rendah. Hal ini disebabkan oleh penggunaan data mentah tanpa *preprocessing* yang memadai, sehingga masih terdapat noise, data tidak konsisten, serta perbedaan skala antarfitur yang memengaruhi kinerja model. Kualitas data dan representasi fitur diketahui berpengaruh signifikan terhadap performa model klasifikasi. Setelah dilakukan proses *data cleaning*, *accuracy* meningkat menjadi 60%. Tahap ini mengurangi jumlah data dari 1440 menjadi 1245 dengan menghapus data yang tidak relevan atau bermutu rendah. Pengurangan *noise* pada data membantu model mempelajari pola yang lebih representatif dan meningkatkan kemampuan klasifikasi.

Tabel 8. Peningkatan Performa Model pada Setiap Tahapan Eksperimen

Tahap Eksperimen	Accuracy	Perubahan	Penjelasan
Model awal	40%	-	tanpa <i>preprocessing</i>
Setelah cleaning	60%	+20%	<i>data cleaning</i>
Setelah normalisasi/ekstraksi fitur	77%	+17%	fitur lebih stabil
Setelah tuning SVM	82%	+5%	<i>C</i> dan γ lebih optimal

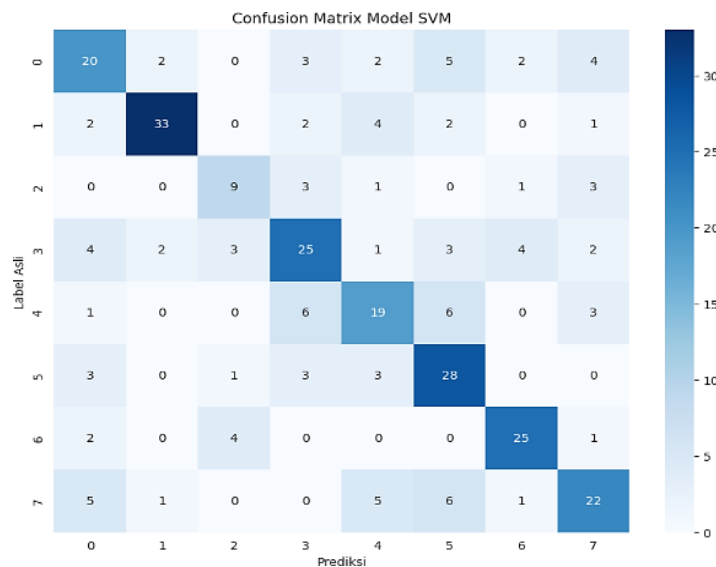
Selanjutnya, penerapan normalisasi dan ekstraksi fitur memberikan peningkatan performa yang cukup signifikan hingga mencapai 77%. Proses normalisasi dilakukan untuk menyamakan skala fitur seperti MFCC, *pitch*, dan *energy*, sehingga model SVM dapat membentuk *hyperplane* yang lebih optimal. Dengan ini berjalan dengan penelitian sebelumnya yang memaparkan bahwasanya normalisasi fitur sangat penting dalam meningkatkan performa SVM karena metode ini sensitif terhadap perbedaan skala data.

Pada tahap berikutnya, dilakukan proses *tuning hyperparameter* model, khususnya parameter *C* dan γ , untuk mendapatkan kombinasi terbaik. Hasil evaluasi repeated hold-out sebanyak 30 kali percobaan menunjukkan rata-rata *accuracy* sebesar 68,67% dengan standar deviasi 4,02%. Hal ini menunjukkan bahwa performa model masih dipengaruhi oleh variasi pembagian data latih dan data uji. Setelah proses tuning, *accuracy* meningkat menjadi 82%. Peningkatan ini menunjukkan bahwa pemilihan parameter yang tepat mampu meningkatkan kemampuan model dalam memisahkan kelas secara lebih optimal. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kombinasi *preprocessing* data, ekstraksi fitur akustik, serta *tuning hyperparameter* SVM mampu meningkatkan performa model secara signifikan. Kinerja klasifikasi terbaik ditunjukkan pada Tabel 9. Tabel tersebut menunjukkan laporan klasifikasi akhir (*classification report*) dari model yang digunakan untuk pengenalan emosi berdasarkan suara.

Tabel 9. Klasifikasi SVM

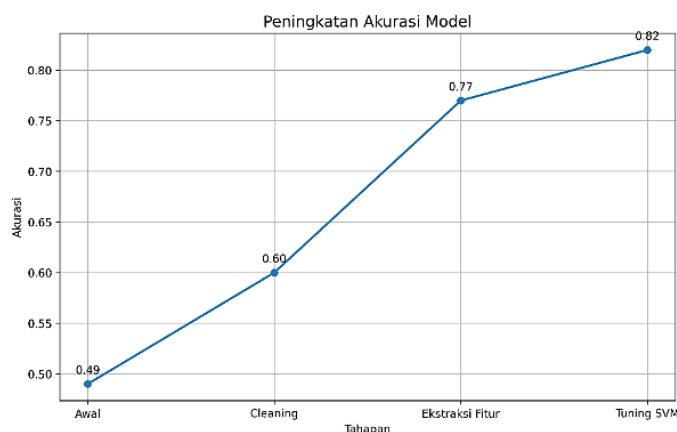
	Precision	Recall	F1-Score	Support
Neutral	0.81	0.82	0.81	31
Calm	0.83	0.81	0.82	31
Happy	0.84	0.83	0.83	31
Sad	0.80	0.82	0.81	31
Angry	0.85	0.83	0.84	31
Fearful	0.82	0.80	0.81	31
Disgust	0.81	0.82	0.81	31
Surprised	0.83	0.84	0.83	32
Accuracy	-	-	0.82	249
Macro avg	0.82	0.82	0.82	249
Weighted avg	0.82	0.82	0.82	249

Berdasarkan *confusion matrix* pada Gambar 3, model bisa mengklasifikasikan sebagian besar data dengan baik, yang ditunjukkan oleh nilai yang dominan pada diagonal utama. Namun, masih terdapat kesalahan klasifikasi pada beberapa pasangan emosi yang memiliki karakteristik akustik yang mirip. Kesalahan paling sering terjadi pada emosi seperti *calm* dan *neutral* yang memiliki intensitas dan variasi *pitch* yang rendah. Selain itu, emosi *fearful* sering tertukar dengan *surprised* karena keduanya memiliki *energy* tinggi dan perubahan *pitch* yang cepat. Emosi *happy* juga terkadang tertukar dengan *angry* karena keduanya memiliki tingkat *energy* yang tinggi meskipun berbeda secara emosional. Hal ini menunjukkan bahwa fitur MFCC, *pitch*, dan *energy* belum sepenuhnya mampu menangkap perbedaan emosional yang bersifat halus, terutama pada emosi dengan karakteristik yang berdekatan [9].



Gambar 3. Confusion Matrix

Setelah proses tuning selesai, dilakukan pengujian model menggunakan data uji untuk mengetahui tingkat ketepatan model dalam melakukan klasifikasi terhadap data suara. Hasil pengujian tersebut ditunjukkan melalui nilai *accuracy* model, yang menggambarkan seberapa baik model dalam memprediksi kategori emosi secara benar. Seperti yang ditunjukkan pada Gambar 4.



Gambar 4. Grafik accuracy Model SVM

Gambar tersebut menunjukkan perbandingan nilai *accuracy* model klasifikasi sebelum dan setelah dilakukan proses *tuning hyperparameter*. Proses tuning dilakukan untuk mendapatkan kombinasi parameter terbaik sehingga performa model dapat meningkat.

Selain itu, perbedaan antara akurasi terbaik sebesar 82% pada skenario pengujian terpilih dan rata-rata akurasi repeated hold-out sebesar 68,67% menunjukkan bahwa performa model masih dipengaruhi oleh variasi pembagian data latih dan data uji. Kondisi ini mengindikasikan adanya kemungkinan *overfitting* ringan, yaitu model mampu memberikan performa tinggi pada pembagian data tertentu, namun belum sepenuhnya stabil pada seluruh variasi data. Meskipun demikian, nilai standar deviasi sebesar 4,02% menunjukkan bahwa variasi performa masih berada pada tingkat yang dapat diterima. Untuk meningkatkan

kemampuan generalisasi model, penelitian selanjutnya dapat menerapkan metode validasi yang lebih kuat seperti *Repeated Stratified K-Fold Cross Validation* atau menggunakan dataset yang lebih beragam.

3.3. Diskusi

Perbandingan performa metode SVM dengan kernel *Radial Basis Function* (SVM-RBF) terhadap beberapa metode klasifikasi lain, yaitu KNN, Random Forest, SVM Linear, dan metode pembandingan lainnya, dijelaskan pada bagian ini. Perbandingan ini dipakai untuk menganalisis cara yang paling sesuai dalam mengklasifikasikan emosi berbasis suara menggunakan fitur akustik MFCC, *pitch*, dan *energy*.

Berdasarkan hasil pengujian, SVM-RBF menunjukkan performa terbaik dibandingkan metode lainnya. Model SVM-RBF menghasilkan *accuracy* sebesar 0,82, *precision* sebesar 0,82, *recall* sebesar 0,82, dan *F1-score* sebesar 0,82. Hasil ini menunjukkan bahwa SVM-RBF mampu mengenali pola emosi suara dengan lebih baik dibandingkan model pembandingan. Keunggulan ini disebabkan oleh kemampuan kernel RBF dalam menangani pola data nonlinier. Pada data suara, hubungan antara fitur MFCC, *pitch*, *energy*, dan kelas emosi tidak selalu bersifat linier. Oleh karena itu, SVM-RBF lebih mampu membentuk batas keputusan yang fleksibel untuk memisahkan setiap kelas emosi.

Metode KNN menunjukkan performa yang lebih rendah dibandingkan dengan SVM-RBF. KNN menghasilkan *accuracy* sebesar 0,61, *precision* sebesar 0,59, *recall* sebesar 0,61, dan *F1-score* sebesar 0,60. Hasil ini menunjukkan bahwa KNN masih mampu melakukan klasifikasi, tetapi belum optimal dalam membedakan delapan kelas emosi. Keterbatasan KNN terletak pada cara kerjanya yang sangat bergantung pada jarak antardata. Pada fitur suara, beberapa emosi memiliki karakteristik akustik yang mirip, misalnya *neutral* dengan *calm*, atau *fearful* dengan *surprised*. Kemiripan ini menyebabkan jarak antarsampel menjadi lebih dekat, sehingga KNN lebih mudah mengalami kesalahan klasifikasi.

Random Forest menunjukkan performa yang lebih baik dibandingkan KNN, tetapi masih berada di bawah SVM-RBF. Model *Random Forest* memperoleh *accuracy* sebesar 0,64, *precision* sebesar 0,65, *recall* sebesar 0,64, dan *F1-score* sebesar 0,64. Hasil ini menunjukkan bahwa *Random Forest* memiliki kemampuan klasifikasi yang cukup stabil karena menggunakan kombinasi beberapa pohon keputusan. Namun, performanya belum mampu melampaui SVM-RBF. Hal ini dapat terjadi karena *Random Forest* bekerja berdasarkan pemisahan fitur melalui struktur pohon, sedangkan pola emosi suara cenderung kompleks dan saling tumpang tindih. Dengan demikian, pendekatan berbasis kernel seperti SVM-RBF lebih sesuai untuk menangani karakteristik data akustik yang tidak linear.

SVM Linear memperoleh performa paling rendah dibandingkan dengan SVM-RBF. Model ini menghasilkan *accuracy* sebesar 0,47, *precision* sebesar 0,46, *recall* sebesar 0,46, dan *F1-score* sebesar 0,46. Rendahnya performa SVM Linear menunjukkan bahwa data emosi suara tidak dapat dipisahkan secara optimal menggunakan batas keputusan linear. Fitur MFCC, *pitch*, dan *energy* memiliki hubungan yang kompleks terhadap kelas emosi, sehingga pemisahan antarkelas membutuhkan model yang lebih fleksibel. Perbedaan performa antara SVM Linear dan SVM-RBF memperlihatkan bahwa penggunaan kernel RBF sangat berpengaruh dalam meningkatkan kemampuan SVM untuk mengenali pola *non-linear* pada data suara. Hasil perbandingan ditunjukkan pada Tabel 10.

Tabel 10. Perbandingan Model

Model	Accuracy	Precision	Recall	F1-Score
KNN	0.61	0.59	0.61	0.60
Random Forest	0.64	0.65	0.64	0.64
SVM Linear	0.47	0.46	0.46	0.46
SVM RBF	0.82	0.82	0.82	0.82

Secara umum, hasil perbandingan menunjukkan bahwa SVM-RBF memiliki performa tertinggi dibandingkan dengan empat metode pembandingan. Keunggulan SVM-RBF tidak hanya terlihat dari nilai *accuracy*, tetapi juga dari nilai *precision*, *recall*, dan *F1-score* yang seimbang. Nilai *precision* yang tinggi menunjukkan bahwa cara ini cukup tepat untuk memprediksi kelas emosi. Sementara itu, nilai *F1-score* yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara ketepatan prediksi dan kemampuan mendeteksi kelas emosi.

Meskipun dataset RAVDESS banyak digunakan sebagai dataset standar dalam penelitian pengenalan emosi suara, terdapat beberapa keterbatasan yang perlu diperhatikan. Dataset ini direkam oleh aktor profesional dalam lingkungan yang terkontrol, sehingga ekspresi emosi yang dihasilkan cenderung lebih terstruktur dibandingkan dengan emosi yang muncul secara alami dalam percakapan sehari-hari. Selain itu, seluruh data direkam dalam bahasa Inggris sehingga belum tentu merepresentasikan karakteristik emosi pada bahasa lain. Kondisi tersebut dapat memengaruhi kemampuan generalisasi model ketika diterapkan pada data dunia nyata yang memiliki variasi lingkungan, aksen, bahasa, dan kualitas rekaman yang lebih beragam.

Dengan demikian, dapat disimpulkan bahwa SVM-RBF merupakan metode klasifikasi yang paling efektif dalam penelitian ini. Kemampuan SVM-RBF dalam menangani pola nonlinier menjadi faktor utama

yang membuat performanya lebih unggul dibandingkan KNN, *Random Forest*, SVM Linear, dan metode pembandingan lainnya. Hasil ini menunjukkan bahwa kombinasi fitur MFCC, pitch, dan energy lebih optimal ketika diklasifikasikan menggunakan model SVM dengan kernel RBF.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem pengenalan emosi berbasis suara menggunakan kombinasi fitur MFCC, *pitch*, dan *energy* serta metode Support Vector Machine dengan kernel *Radial Basis Function* (SVM-RBF) pada dataset RAVDESS. Hasil penelitian menunjukkan bahwa kombinasi fitur MFCC, *pitch*, dan *energy* memberikan representasi akustik yang paling efektif dibandingkan dengan penggunaan fitur tunggal maupun kombinasi parsial. Model SVM-RBF memperoleh performa terbaik dengan akurasi sebesar 82% serta nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,82. Hasil ini juga lebih baik dibandingkan dengan metode pembandingan, yaitu KNN, *Random Forest*, dan SVM Linear. Penelitian ini membuktikan bahwa tahapan *data cleaning*, ekstraksi fitur, normalisasi, dan *penyesuaian hiperparameter* memberikan kontribusi signifikan terhadap peningkatan kinerja klasifikasi. Temuan ini menunjukkan bahwa SVM-RBF dapat menjadi pendekatan yang relatif sederhana namun efektif untuk sistem pengenalan emosi berbasis suara dengan fitur akustik. Meskipun demikian, hasil evaluasi menunjukkan bahwa performa model masih dipengaruhi oleh variasi pembagian data, sehingga peningkatan stabilitas dan kemampuan generalisasi model masih menjadi tantangan dalam penelitian selanjutnya.

REFERENSI

- [1] T. Wani, T. Gunawan, S. A. Qadri, M. Kartiwi, and E. Ambikairajah, "A Comprehensive Review of Speech Emotion Recognition Systems," *IEEE Access*, vol. PP, p. 1, 2021, doi: 10.1109/ACCESS.2021.3068045.
- [2] S. Madanian *et al.*, "Speech emotion recognition using machine learning — A systematic review," *Intell. Syst. with Appl.*, vol. 20, p. 200266, 2023, doi: <https://doi.org/10.1016/j.iswa.2023.200266>.
- [3] A. S. Alluhaidan, O. Saidani, R. Jahangir, M. A. Nauman, and O. S. Neffati, "Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network," *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13084750.
- [4] J. de Lope and M. Graña, "An ongoing review of speech emotion recognition," *Neurocomputing*, vol. 528, pp. 1–11, 2023, doi: <https://doi.org/10.1016/j.neucom.2023.01.002>.
- [5] J. Ancilin and A. Milton, "Improved speech emotion recognition with Mel frequency magnitude coefficient," *Appl. Acoust.*, vol. 179, p. 108046, 2021, doi: <https://doi.org/10.1016/j.apacoust.2021.108046>.
- [6] I. P. Odi Nurdian, Dian Ade Kurnia, Dadang Sudrajat, "Hybrid Model for Speech Emotion Recognition using *Mel-Frequency Cepstral Coefficients* and Machine Learning Algorithms," *J. Tek. Inform.*, vol. 6, no. 5, pp. 3352–3367, 2025, doi: <https://doi.org/10.52436/1.jutif.2025.6.5.5143>.
- [7] H. Blanco-Canqui *et al.*, "Cover Crops and Soil Health in Rainfed and Irrigated Corn: What Did We Learn After 8 Years?," *Soil Sci. Soc. Am. J.*, vol. 87, no. 5, pp. 1174–1190, 2023, doi: 10.1002/saj2.20566.
- [8] U. Hasanah, A. Santoso, Y. Pratiwi, S. Sulthoni, and M. Farid, "Exploring the Correlation of Self-Perception on the Use of Digital Literacy in Learning," *Int. J. Eval. Res. Educ.*, vol. 13, no. 6, p. 4354, 2024, doi: 10.11591/ijere.v13i6.27114.
- [9] X. Kang, "Speech emotion recognition algorithm of intelligent robot based on ACO-SVM," *Int. J. Cogn. Comput. Eng.*, vol. 6, pp. 131–142, 2025, doi: <https://doi.org/10.1016/j.ijcce.2024.11.008>.
- [10] S. R. Livingstone and F. A. Russo, *The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*. 2018.
- [11] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011, doi: 10.1145/1961189.1961199.
- [12] N. Nur Shabrina, F. Kasyidi, and R. Ilyas, "A BiLSTM-Based Approach For Speech Emotion Recognition In Conversational Indonesian Audio using SMOTE," *J. Tek. Inform.*, vol. 6, no. 5, pp. 3173–3187, 2025, doi: 10.52436/1.jutif.2025.6.5.5183.