

## METODOLOGI

## Akses terbuka



# Ekstraksi fitur paralinguistik dan spektral untuk klasifikasi emosi ucapan menggunakan teknik pembelajaran mesin

Tong Liu<sup>1</sup> dan Xiaochen Yuan<sup>1\*</sup>

## Abstrak

Emosi memainkan peran dominan dalam ucapan. Ucapan yang sama dengan emosi yang berbeda dapat menimbulkan makna yang sangat berbeda. Kemampuan menampilkan berbagai emosi saat berbicara juga merupakan salah satu ciri khas manusia. Dalam hal ini, tren teknologi mengembangkan algoritma klasifikasi emosi ucapan yang canggih dalam rangka meningkatkan interaksi antara komputer dan manusia. Makalah ini mengusulkan pendekatan klasifikasi emosi ujaran berdasarkan ekstraksi fitur paralinguistik dan spektral. Koefisien cepstral frekuensi Mel (MFCC) diekstraksi sebagai fitur spektral, dan openSMILE digunakan untuk mengekstrak fitur paralinguistik. Teknik pembelajaran mesin pengklasifikasi perceptron multi-layer dan mesin vektor pendukung masing-masing diterapkan ke dalam fitur yang diekstraksi untuk klasifikasi emosi ucapan. Kami telah melakukan percobaan pada database Berlin untuk mengevaluasi kinerja pendekatan yang diusulkan. Hasil eksperimen menunjukkan bahwa pendekatan yang diusulkan mencapai kinerja yang memuaskan. Perbandingan dilakukan masing-masing dalam kondisi bersih dan kondisi bising, dan hasilnya menunjukkan kinerja yang lebih baik dari skema yang diusulkan.

**Kata Kunci** Klasifikasi emosi ucapan, Fitur paralinguistik, Fitur spektral, Mesin vektor dukungan, Pengklasifikasi perceptron multi-layer

## 1. Perkenalan

Dalam beberapa tahun terakhir, meningkatnya permintaan akan kecerdasan buatan telah menjadikan program interaksi manusia-komputer sebagai aspek penting dalam teknologi komputer. Dalam hal ini, emosi manusia, khususnya deteksi emosi ucapan, telah mendapat perhatian besar. Ucapan yang sama dengan emosi yang berbeda dapat menimbulkan makna yang sangat berbeda. Kemampuan menampilkan berbagai emosi saat berbicara juga merupakan salah satu ciri khas manusia. Oleh karena itu, tren teknologi mengembangkan sistem pengenalan emosi ucapan yang canggih dalam rangka meningkatkan interaksi antara komputer dan manusia, dan dengan demikian klasifikasi/pengenalan emosi secara bertahap menjadi

aplikasi penting yang tidak dapat disangkal dalam pemrosesan sinyal suara dan interaksi manusia-komputer [1-4].

Sebenarnya klasifikasi emosi didasarkan pada berbagai model psikologi, yang mengklasifikasikan emosi menurut aturan dan prinsip tertentu, membantu kita lebih memahami dan menjelaskan emosi. Berikut beberapa model klasifikasi emosi yang umum digunakan. Yang paling mendasar adalah Enam Model Emosi Dasar [5], yang mengkategorikan emosi menjadi enam jenis: marah, jijik, takut, bahagia, sedih, dan terkejut. Enam Model Emosi Dasar (SBEM) adalah salah satu model yang paling terkenal dan banyak digunakan dalam klasifikasi emosi dan telah diterapkan di berbagai bidang seperti psikologi, ilmu saraf, dan grafik komputer.

Namun, model ini dibatasi oleh penyederhanaan yang berlebihan, karena emosi bersifat kompleks dan beberapa emosi tidak dapat diklasifikasikan menggunakan model ini. Oleh karena itu, diusulkan model emosi 2D, Circumplex Model of Affect [6], di mana emosi disajikan oleh gairah dan valensi. Tere menjadi emosional

\*Korespondensi:  
Xiaochenyuan

xcyuan@mpu.edu.mo

<sup>1</sup> Fakultas Ilmu Terapan, Universitas Politeknik Macao, Macao, Cina

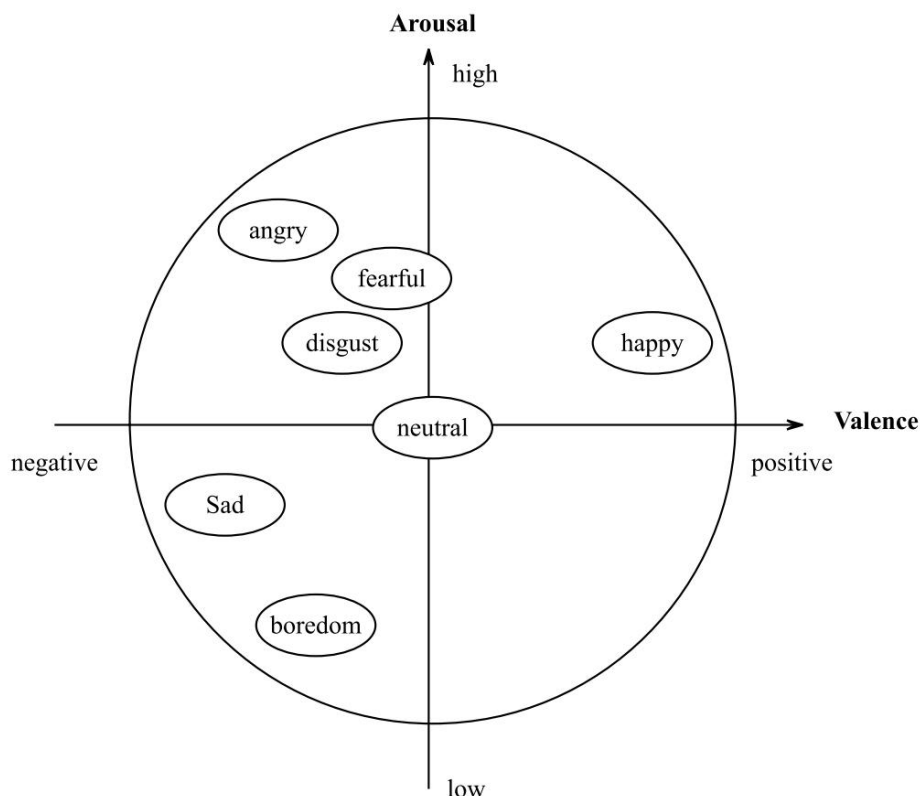
gairah mengacu pada intensitas atau tingkat emosi, dan valensi emosional mengacu pada respon fisiologis dan kecenderungan perilaku emosi. Gambar 1 menampilkan Circumplex Model of Affect yang diterapkan pada kelas emosi Berlin Dataset [7]. Atas dasar ini, ada model klasifikasi emosi umum lainnya yang disebut model emosi tiga faktor [8]. Selain gairah dan valensi emosional, model ini menambahkan signifikansi emosional, yang dapat menjelaskan kinerja dan perubahan emosi dengan lebih baik dalam konteks yang berbeda.

Secara umum, sistem klasifikasi emosi ujaran terdiri dari dua bagian: ekstraksi fitur dan klasifikasi fitur. Ekstraksi fitur berkaitan dengan ekstraksi fitur ucapan yang terkait dengan emosi menggunakan beberapa alat ekstraktor. Dalam klasifikasi fitur, ini memproses fitur yang diekstraksi untuk melatih model menggunakan pengklasifikasi untuk memprediksi kategori emosi. Dengan demikian, pemilihan fitur emosi menjadi poin kuncinya. Berbagai fitur ucapan telah digunakan dalam klasifikasi emosi [9, 10]. Telah dilaporkan ada tiga kelas ciri-ciri tradisional berdasarkan cara manusia berbicara: ciri-ciri prosodik [11], ciri-ciri suku kata [12], dan ciri-ciri spektral [13, 14]. Ciri prosodik disebut juga ciri suprasegmental atau ciri paralinguistik [11]. Hal ini terutama dikendalikan oleh nada, energi, dan frekuensi bicara [15]. Fitur suku kata

sebagian besar terkait dengan kualitas suara. Fitur spektral dianggap sebagai refleksi hubungan antara jalur suara yang bergerak dan produksi suara. Pada sebagian besar waktu, koefisien cepstrum prediksi linier (LPCC) [16] dan koefisien cepstral frekuensi skala Mel (MFCC) [17] mewakili genre fitur ini. Banyak penelitian telah dilakukan dengan menggunakan berbagai fitur.

Nwe dkk. [18] mengusulkan klasifikasi emosi ucapan berdasarkan model Markov menggunakan koefisien daya frekuensi log waktu singkat (LFPC). Mereka mengusulkan metode klasifikasi emosi ucapan yang independen terhadap teks, yang memanfaatkan LFPC untuk merepresentasikan sinyal ucapan dan menggunakan model Markov tersembunyi diskrit (HMM) sebagai pengklasifikasinya. Kinerja yang diperoleh dari metode mereka telah dibandingkan dengan fitur LPCC [16] dan MFCC yang umum digunakan dalam sistem klasifikasi ucapan.

Wong dkk. [2] melakukan penelitian tentang klasifikasi emosi ucapan dengan menggunakan parameter Fourier yang diusulkan. Parameter ini berkaitan erat dengan kualitas ucapan dan menggunakan konten persepsinya dengan perbedaan tingkat pertama dan kedua untuk klasifikasi emosi ucapan yang tidak bergantung pada pembicara. Karena perbedaan karakteristik bicara antara anak-anak dan orang dewasa, suara anak-anak biasanya memiliki nada yang lebih tinggi dan kecepatan bicara yang lebih cepat, Albu dkk. [19] memanfaatkan MFCC dan beberapa parameter untuk mendapatkan



**Gambar 1** Memetakan kelas emosi Kumpulan Data Berlin ke dalam Model Pengaruh Circumplex [7]

vektor fitur untuk masukan jaringan saraf. Parameter tambahan ini mencakup energi jangka pendek, laju zero-crossing, roll-of spektral, dan pusat massa spektral. Rata-rata dan deviasi standar fitur-fitur ini kemudian dihitung dan digabungkan untuk membuat vektor fitur akhir untuk sinyal ucapan.

Dalam aspek metode klasifikasi, berbagai pilihan dapat diambil untuk sistem emosi bicara [20, 21]. Khususnya, mesin pembelajaran ekstrim (ELM) [20], K-near-est neighbour (KNN) [21], dan mesin vektor dukungan (SVM) [22] semuanya merupakan pengklasifikasi yang identik dan banyak digunakan.

Setiap classifier memiliki propertinya sendiri yang sesuai untuk jenis aplikasi tertentu. Untuk meningkatkan kinerja klasifikasi, berbagai metode pemilihan fitur bersifat opsional. Jaringan spektral dalam (DNN) juga telah digunakan dalam klasifikasi emosi ucapan.

Seperti yang disajikan dalam [23], distribusi probabilitas keadaan emosi untuk setiap segmen ucapan dihasilkan menggunakan DNN, dan fitur tingkat ucapan dibangun dari distribusi probabilitas tingkat segmen. Sedangkan untuk klasifikasi emosi anak, Albu dkk. [19] menguji kinerja menggunakan jaringan radial basis function (RBF), ELM, dan online sequence ELM (OS-ELM). Hasil eksperimen menunjukkan bahwa jaringan RBF mencapai kinerja yang lebih unggul dibandingkan ELM.

Namun, meskipun banyak penelitian telah dilakukan selama bertahun-tahun, klasifikasi menggunakan fitur-fitur ini dalam lingkungan yang bising masih menjadi masalah besar. Kinerja klasifikasi akan dihancurkan dengan sinyal-sinyal yang terdistorsi kebisingan. Oleh karena itu, dalam penelitian ini, fitur paralinguistik tradisional yang disebutkan di atas dan fitur spektral diekstraksi untuk memproses prosedur klasifikasi. Fitur spektral disajikan oleh model komputasi yang menghasilkan serangkaian respons frekuensi karakteristik tertentu dari suatu ucapan melalui serat saraf pendengaran. Koefisien MFCC diekstraksi sebagai fitur spektral. Sementara itu, kami mengusulkan untuk mengekstrak fitur paralinguistik tradisional yang disajikan oleh rangkaian tantangan paralinguistik INTERSPEECH 2013 [24] dari klasifikasi emosi ucapan untuk ucapan, menggunakan perangkat OpenSMILE [25]. Berbagai macam fitur paralinguistik dan fitur prosodik disertakan dalam set ini.

Te INTERSPEECH 2013 berisi 6373 fitur, LLD termasuk energi, spektrum, cepstrum, suara, log harmonic noise ratio (HNR), harmoni spektral, dan kejernihan spektral psi-koakustik. Untuk mengekstrak fitur, dilakukan prosedur sebagai berikut: mengekstraksi deskriptor tingkat rendah, menambahkan jendela Hamming, menghaluskan fitur, menambahkan koefisien, dan menerapkan fungsi pada masing-masing deskriptor.

Setelah mengekstraksi fitur-fiturnya, kami menggunakan multi-layer perceptron classifier (MLP classifier) [26] dan SVM masing-masing untuk klasifikasi emosi ucapan.

Basis data Berlin digunakan untuk mengevaluasi kinerja, dan hasil eksperimen menunjukkan bahwa pendekatan yang diusulkan mencapai kinerja yang baik dalam kondisi yang berbeda dan berkinerja lebih baik dibandingkan pekerjaan terkait dalam hal berbagai metrik evaluasi. Struktur makalah ini disusun sebagai berikut: Bagian 2 menjelaskan pendekatan yang diusulkan secara rinci, Bagian 3 mengilustrasikan hasil eksperimen, dan Bagian 4 menyimpulkan makalah dan memberikan karya masa depan.

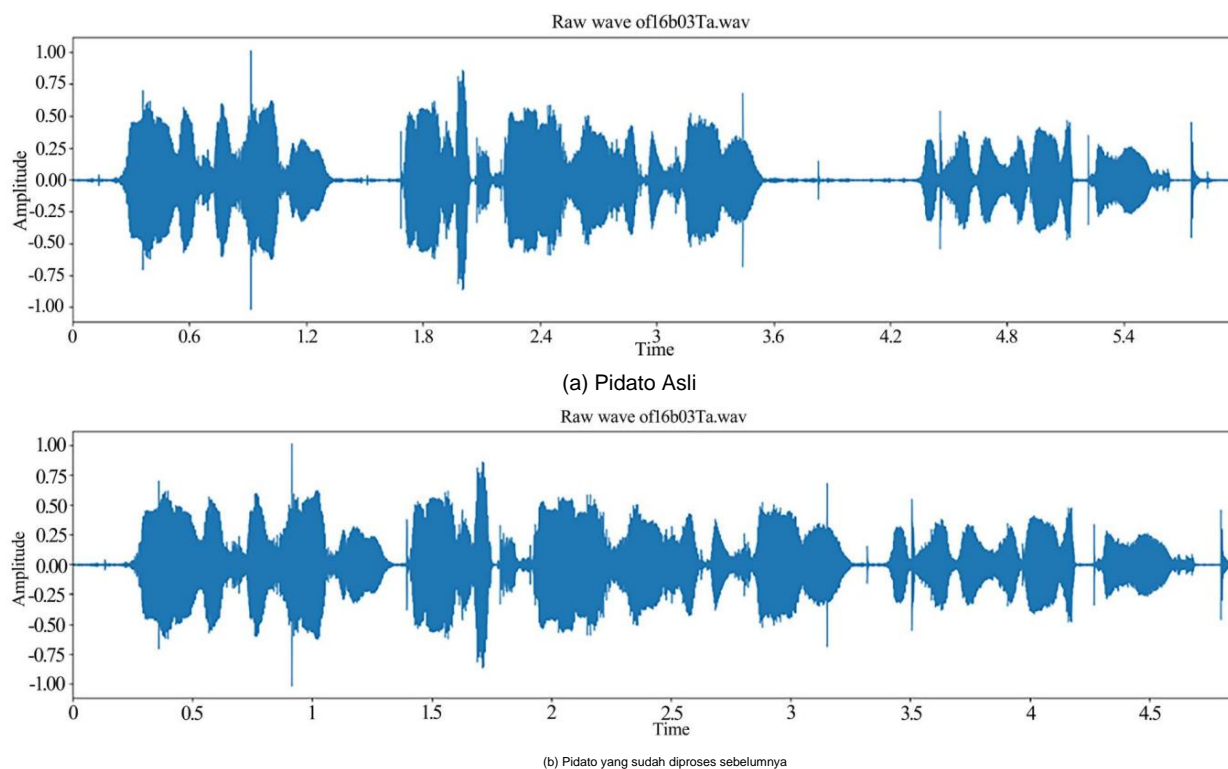
## 2 Usulan pendekatan klasifikasi emosi ucapan

Penting untuk disebutkan bahwa database pengujian memiliki segmen diam; oleh karena itu, data tersebut perlu diolah terlebih dahulu agar sampel tuturan tersebut sesuai dengan metode penelitian ini; Sementara itu, pra-pemrosesan juga memiliki kemampuan untuk meningkatkan akurasi hasil klasifikasi. Dalam penelitian ini, sampel kecepatan bicara akan diekstraksi oleh Librosa [27], yang merupakan kotak alat yang disediakan oleh Python, dan terutama digunakan dalam memproses audio atau menganalisis audio. Segmen diam akan dihapus oleh kotak alat IPython, yang menggunakan laju sampel untuk menentukan laju yang berguna dan menghapus laju yang tidak berguna. Gambar 2 menunjukkan contoh spektrum ucapan sebelum dan sesudah pra-pemrosesan. (a) menunjukkan spektrum sampel ucapan asli, dan (b) menunjukkan sampel setelah pra-pemrosesan.

Setelah prapemrosesan, kami masing-masing menggunakan OpenSMILE dan MFCC untuk ekstraksi fitur. OpenSMILE adalah perangkat lunak sumber terbuka untuk secara otomatis mengekstraksi fitur dari sinyal audio dan mengklasifikasikan sinyal ucapan dan musik. Rangkaian fitur standar yang digunakan dalam penelitian ini adalah rangkaian "emobase" openSMILE/openEAR, yang memiliki 998 fitur akustik untuk klasifikasi emosi, termasuk berbagai deskriptor tingkat rendah (LLD): intensitas, kenyaringan, 12 MFCC, nada (F0), probabilitas yang disuarakan, amplop F0, frekuensi spektral 8 garis, laju persilangan nol, dan koefisien regresi delta.

MFCC adalah fitur yang banyak digunakan dalam klasifikasi ucapan dan pembicara otomatis karena cocok untuk memahami manusia dan frekuensi di mana manusia berbicara.

Ekstraksi fitur MFCC terdiri dari dua langkah utama: analisis frekuensi mel dan analisis cepstral. MFCC adalah seperangkat koefisien kunci yang digunakan untuk menetapkan Mel cepstral. Dari segmen sinyal audio, kita dapat memperoleh sekumpulan cepstrum yang cukup untuk mewakili sinyal audio. Berbeda dengan cepstrum pada umumnya, pita frekuensi pada Mel cepstrum tersebar merata pada skala Mel, pita frekuensi seperti itu akan lebih linier dari yang biasa kita lihat. Metode representasi cepstral lebih dekat dengan sistem pendengaran nonlinier manusia. Oleh karena itu, dalam metode ini kami memilih MFCC sebagai metode ekstraksi fitur.



Gambar 2 Contoh spektrum ucapan setelah pra-pemrosesan

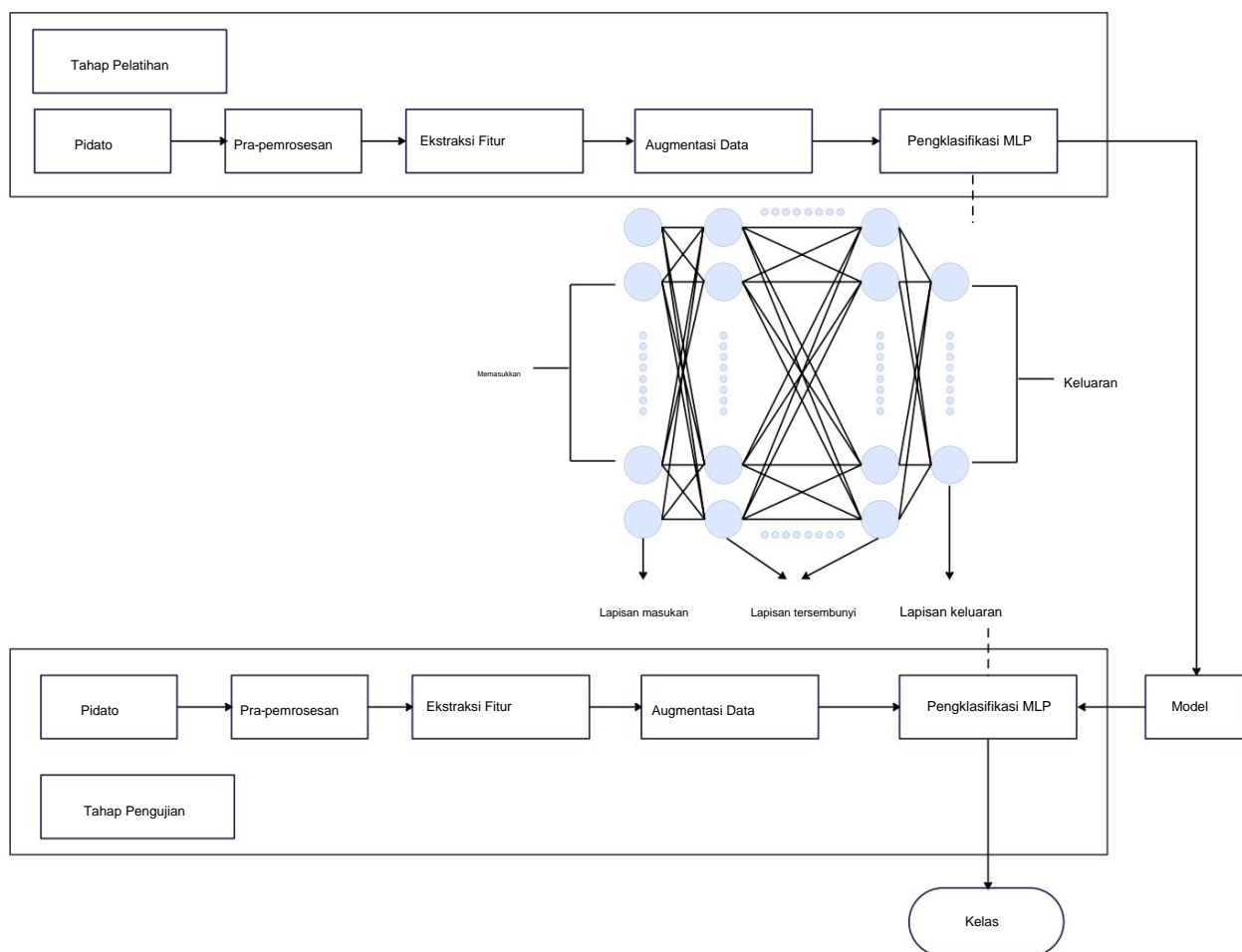
## 2.1 Klasifikasi menggunakan MLP classifier

Gambar 3 menunjukkan diagram skema SEC-MLP yang diusulkan. Pengklasifikasi MLP adalah jaringan yang terdiri dari perceptron. Strukturnya terdiri dari lapisan masukan **Lin**,  $n$  lapisan tersembunyi **Lh**, dan lapisan keluaran **Lout**. Setiap lapisan terdiri dari beberapa neuron, dan setiap neuron dalam suatu lapisan terhubung sepenuhnya dengan setiap neuron pada lapisan yang berdekatan (28). Te **Lin** adalah lapisan pertama jaringan saraf, yang menerima data masukan dan meneruskannya ke lapisan berikutnya. Dan setiap neuron berhubungan dengan satu fitur masukan; oleh karena itu, jumlah neuron **Nin** di **Lin** bergantung pada jumlah fitur masukan. Lapisan keluaran **Lout** digunakan untuk membuat prediksi untuk masukan yang diberikan. Setiap neuron di **Lout** mewakili sebuah kelas dan mengeluarkan probabilitas untuk menjadi bagian dari kelas yang sesuai. Secara umum, jumlah neuron di **Lout** sama dengan jumlah kelas dalam masalah klasifikasi. Jika inputnya perlu diklasifikasikan ke dalam **Nout** kelas, maka lapisan keluaran harus memiliki neuron **Nout**. Lapisan yang ada di antara lapisan masukan dan lapisan keluaran disebut lapisan tersembunyi **Lh**. Setiap **Lh** jaringan saraf menerima sinyal dari keluaran lapisan sebelumnya dan meneruskannya ke lapisan berikutnya sebagai masukan melalui koneksi berbobot. Setiap **Lh** juga memiliki neuron **Nh**, dan jumlah  $n$  dan **Nh** dapat disesuaikan dengan kebutuhan tugas. Untuk tugas pengenalan emosi ucapan, itu

fitur masukan tipikal dengan dimensi tinggi, sedangkan jumlah kelas keluaran relatif kecil. Oleh karena itu, memilih jumlah neuron yang lebih kecil dapat membantu mengurangi kompleksitas jaringan saraf dan mengurangi waktu pelatihan dan penggunaan memori model [29, 30].

Dalam tulisan ini, kami menetapkan jumlah neuron di setiap **Lh** hingga 300. Mengingat  $k$  sampel pelatihan, MLP dengan dimensi fitur masukan **Nin**,  $n$  lapisan tersembunyi dengan **Nh** neuron di setiap lapisan tersembunyi, dan neuron keluaran **Nout**, kompleksitas waktu untuk klasifikasi dalam satu iterasi adalah  $O(k \times \text{Nin} \times \text{Nh} \times \text{Nout})$ .

Selain itu, augmentasi data digunakan sebelum menerapkan pengklasifikasi MLP. Augmentasi data adalah serangkaian teknik untuk meningkatkan jumlah data secara artifisial dengan menghasilkan titik data baru dari data yang sudah ada. Hal ini termasuk membuat perubahan kecil pada data atau menggunakan model pembelajaran mendalam untuk menghasilkan titik data baru. Hal ini berguna untuk meningkatkan performa dan hasil model pembelajaran mesin dengan membentuk contoh baru dan berbeda untuk melatih kumpulan data. Ada banyak metode berbeda untuk mencapai augmentasi data, seperti padding, cropping, penghapusan acak, menambahkan white noise dan penyetelan nada, dan sebagainya. Dalam pekerjaan ini, kami menggunakan lima augmentasi data, termasuk menambahkan white noise, penyetelan nada, pergeseran acak, penyetelan speed dan nada, dan peregangannya.



**Gambar 3** Diagram alir klasifikasi emosi tuturan menggunakan MLP classifier (SEC-MLP)

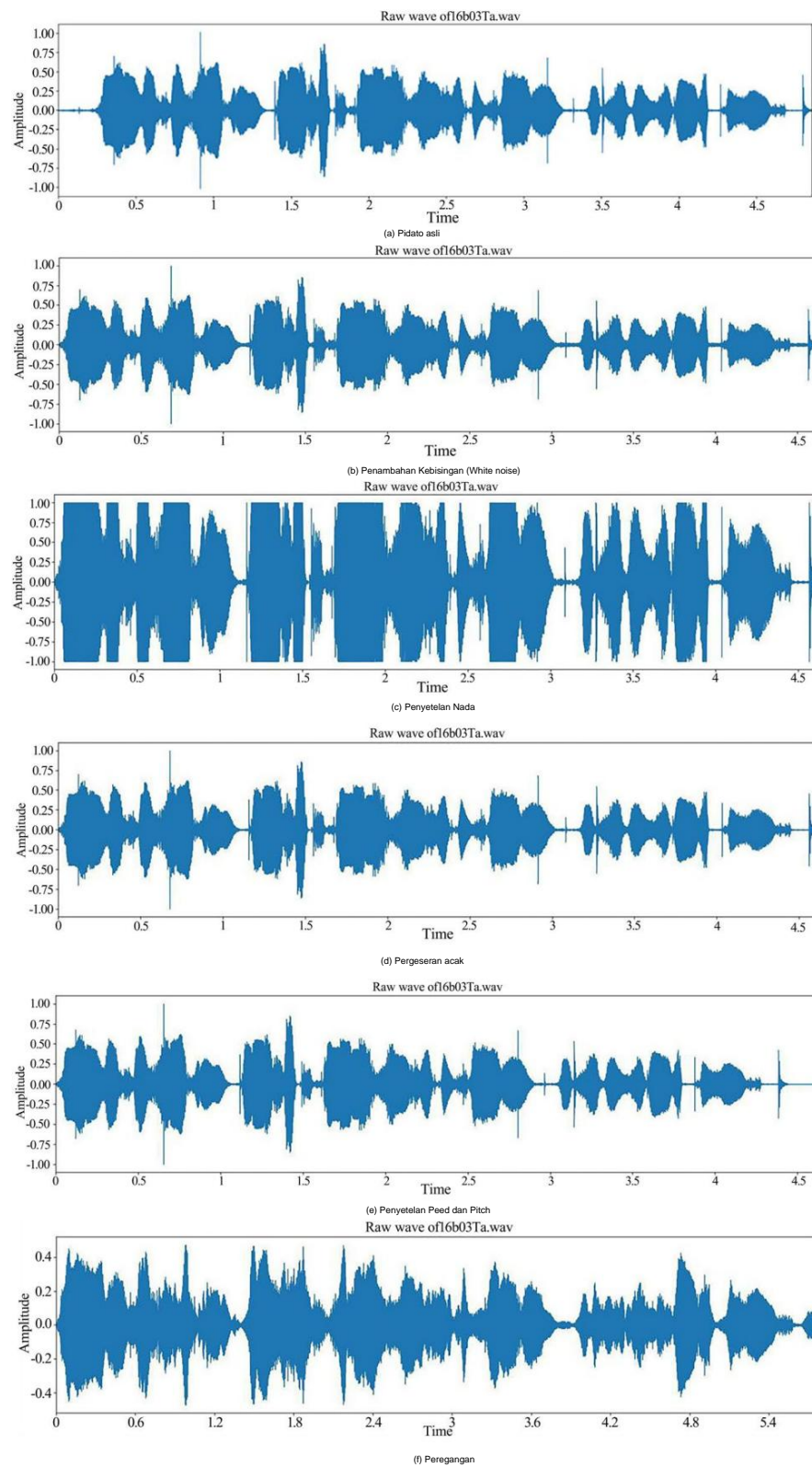
suara. Gambar 4 menunjukkan contoh ucapan dan spektrum setelah setiap penambahan data.

Dalam SEC-MLP yang diusulkan, kami menggunakan jaringan spektral pengklasifikasi MLP yang terhubung penuh untuk klasifikasi. Pengklasifikasi MLP bergantung pada jaringan spektral yang mendasarinya untuk melakukan tugas klasifikasi. Pengklasifikasi MLP dibuat untuk melatih kumpulan data yang diberikan. Fase pelatihan memungkinkan pengklasifikasi MLP mempelajari korelasi antara kumpulan masukan dan keluaran. Selama pelatihan, pengklasifikasi MLP menyesuaikan parameter model seperti bobot dan bias untuk meminimalkan kesalahan. Pengklasifikasi MLP menggunakan propagasi mundur untuk membuat penyesuaian bobot dan bias relatif terhadap kesalahan. Selama fase implementasi, ukuran batch ditetapkan sebesar 256, dan pengoptimal Adam diadopsi bersama dengan pendekatan kecepatan pembelajaran adaptif. Kecepatan pembelajaran diinisialisasi ke 0,001. Jika kecepatan pembelajaran tidak dapat diturunkan selama dua periode berturut-turut, kecepatan pembelajaran saat ini dibagi 5.

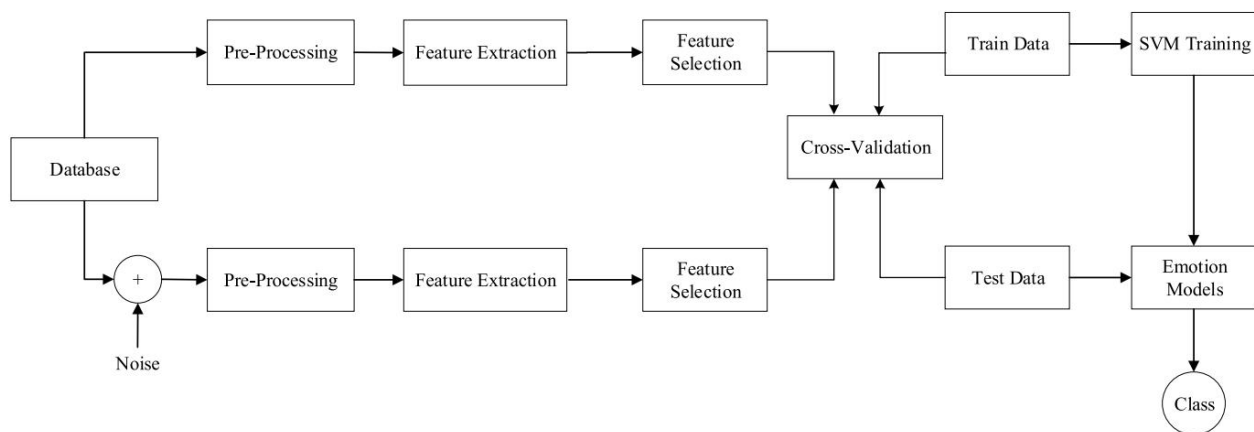
## 2.2 Klasifikasi menggunakan SVM

Gambar 5 menunjukkan empat grafik SEC-SVM yang diusulkan. Dengan fitur yang diekstraksi seperti yang dijelaskan di Bagian 2.1, kami menerapkan algoritma pemilihan fitur untuk meningkatkan kinerja. Untuk memilih fitur, berbagai subset dihasilkan dari semua fitur, dan setiap subset akan disaring dengan algoritma pembelajaran dan dievaluasi berdasarkan kinerjanya. Ini adalah proses memilih subset yang sangat relevan dalam berbagai fitur. Tujuan dari algoritma pemilihan fitur adalah untuk menemukan data yang lebih relevan sehingga kinerja dapat ditingkatkan dengan menghapus fitur-fitur yang kurang relevan. Dengan pemilihan fitur, visualisasi data dan pemahaman data dapat ditingkatkan, kebutuhan pengukuran dan penyimpanan dapat dikurangi, waktu pelatihan, pengujian, dan pemanfaatan dapat dikurangi, efisiensi operasi dapat ditingkatkan, dan pada akhirnya kinerja klasifikasi dapat ditingkatkan. Dalam pekerjaan ini, kinerja masing-masing algoritma dievaluasi dengan hasil rata-rata

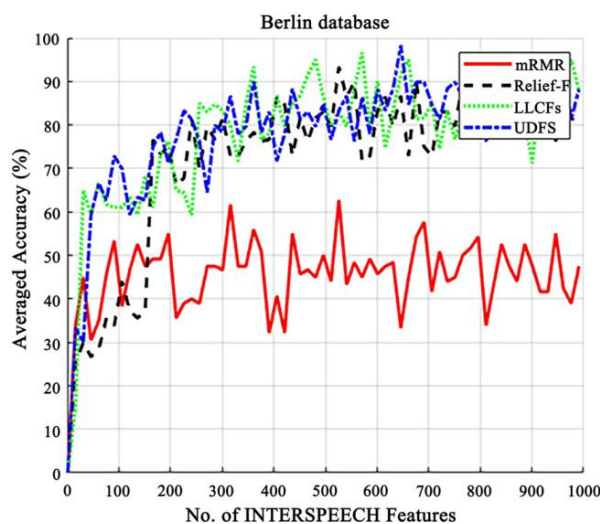




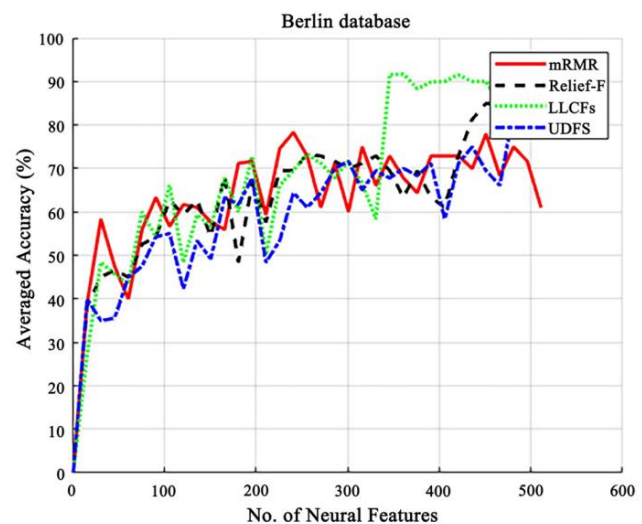
**Gambar 4** Spektrum ucapan setelah augmentasi data. **a** Pidato orisinal, **b** penambahan derau, **c** penyetelan nada, **d** peralihan acak, penyetelan **kecepatan** dan nada, dan **f** peregangan



**Gambar 5** Diagram alir klasifikasi emosi tuturan menggunakan SVM (SEC-SVM)



**Gambar 6** Kinerja algoritma pemilihan fitur yang berbeda

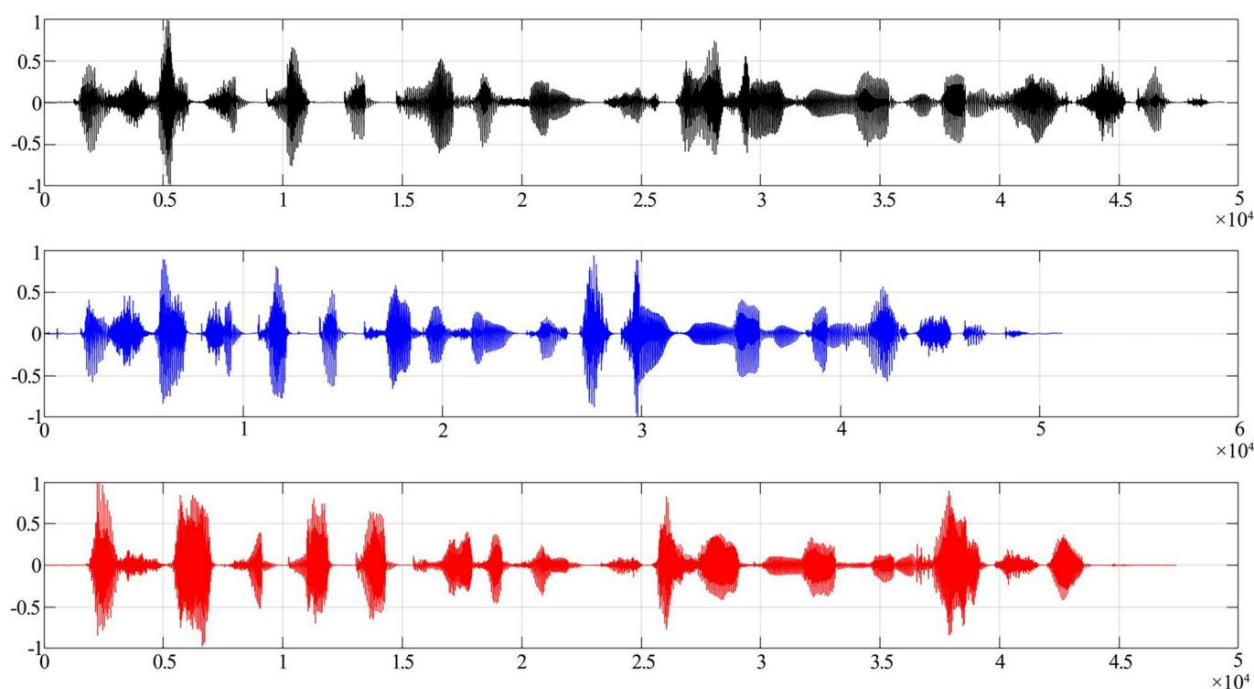


akurasi menggunakan metode pemilihan fitur yang dipilih. Dua kategori algoritma seleksi digunakan untuk evaluasi: algoritma yang diawasi dan algoritma yang tidak diawasi [31]. Pembelajaran yang diawasi adalah tugas pembelajaran mesin untuk mempelajari fungsi yang memetakan masukan ke keluaran berdasarkan contoh pasangan masukan-keluaran [32]. Pembelajaran tanpa pengawasan adalah memodelkan pola tersembunyi atau struktur mendasar dalam data masukan tertentu untuk mempelajari data tersebut. Algoritma yang diawasi adalah pemilihan fitur relevansi maksimum redundansi minimum (mRMR) [33] dan Relief-F [34]. Algoritma yang tidak diawasi adalah pemilihan fitur pengelompokan berbasis pembelajaran lokal (LLCFs) [35] dan pemilihan fitur diskriminatif yang diatur norma L2,1 tanpa pengawasan (UDFS) [36]. Kami menghitung rata-rata ACC dari empat metode pemilihan fitur masing-masing dan menunjukkan hasilnya pada Gambar. 6, di mana

sumbu horizontal menunjukkan jumlah fitur INTERSPEECH yang diekstraksi oleh OpenSMILE. Dapat dilihat bahwa Relief-F, LLCFs, LLCFs, dan UDFS dapat mencapai kinerja yang lebih baik dibandingkan mRMR.

Dalam SEC-SVM, pengklasifikasi SVM diterapkan untuk klasifikasi emosi ucapan. Algoritme pelatihan SVM membangun model yang menetapkan contoh pengujian ke satu kategori atau lainnya. Pertama, label pelatihan, karakteristik pelatihan, label pengujian, dan karakteristik pengujian dihasilkan. Selanjutnya model SVM dibangun dengan menggunakan label dan karakteristik pelatihan di atas. Koefisien konstanta dan gamma terbaik untuk membangun model pelatihan dihitung dengan pencarian grid. Terakhir, SVM-predict mendefinisikan cara memprediksi label dan akurasi prediksi.

Algoritma validasi silang digunakan dalam SEC-SVM. Ide utama validasi silang adalah membagi kumpulan data asli



**Gambar 7** Contoh spektrum diambil dari database Berlin

menjadi beberapa bagian; beberapa di antaranya dikirim ke metode klasifikasi sebagai data pelatihan dan yang lainnya mewakili data pengujian. Proses ini memiliki kemampuan untuk mengevaluasi kinerja masalah klasifikasi dan fags seperti overfitting. Dalam penelitian ini, validasi silang sepuluh kali lipat, yang merupakan metode validasi silang tidak menyeluruh, digunakan untuk membagi kumpulan data menjadi bagian pelatihan dan pengujian.

Validasi silang sepuluh kali lipat memisahkan kumpulan data menjadi sepuluh subkumpulan yang kira-kira sama. Sembilan himpunan bagian ditetapkan sebagai data latih dan satu himpunan bagian sisanya ditetapkan sebagai data uji. Proses ini akan diulang sepuluh kali untuk memastikan setiap bagian dari sepuluh subset telah digunakan sebagai data kereta setidaknya satu kali. Data yang didistribusikan kemudian dikirim untuk penggunaan klasifikasi. Seluruh prosedur akan diulang sebanyak 20 kali. Mengingat  $k$  sampel pelatihan dengan  $m$  fitur, kompleksitas waktu untuk SVM adalah  $O(k \times m^2)$ .

Gambar 6 berisi empat grafik, masing-masing menunjukkan akurasi rata-rata yang diperoleh dengan menggunakan empat metode pemilihan fitur yang disebutkan. Hasilnya diperoleh dengan menyaring seluruh kumpulan fitur dengan empat algoritma pemilihan fitur dan kemudian fitur yang disaring dikirim ke SVM untuk klasifikasi. Sumbu  $x$  pada grafik mewakili jumlah fitur, dan sumbu  $y$  adalah akurasi rata-rata. Tujuan dari pengujian ini adalah untuk mengetahui algoritma pemilihan fitur terbaik dan jumlah fitur yang difilter sehingga dapat memperoleh rata-rata akurasi klasifikasi umur tertinggi. Seperti yang disajikan dalam

Gambar tersebut, kinerja akurasi rata-rata terbaik untuk fitur spektral dicapai oleh algoritma UDFS dan kinerja akurasi rata-rata terbaik untuk fitur INTERSPEECH dicapai oleh algoritma LLCFs. Dengan menggunakan algoritma UDFS untuk fitur spektral dan algoritma LLCFs untuk fitur INTERSPEECH, fitur terbaik dipilih dan dikirim ke SVM untuk klasifikasi.

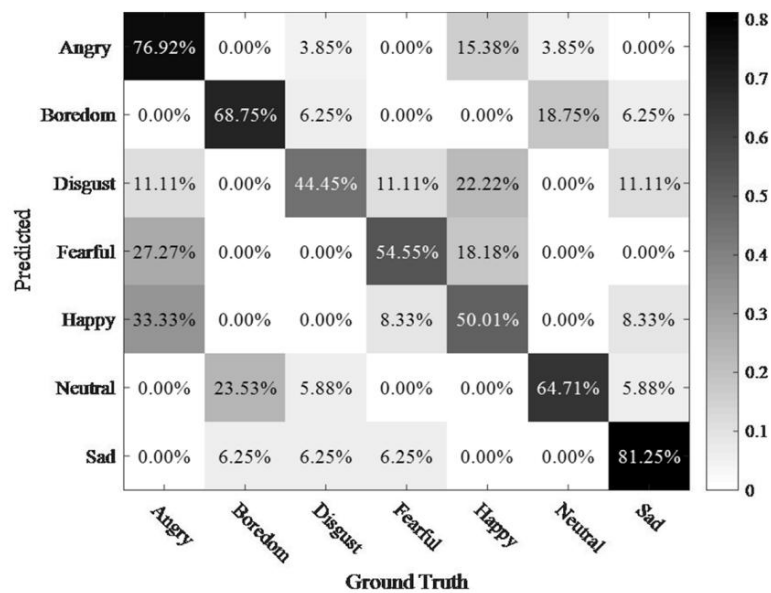
### 3 Hasil percobaan dan pembahasan

Dalam penelitian ini, database terkenal Berlin [37] digunakan untuk menguji kinerja metode yang diusulkan.

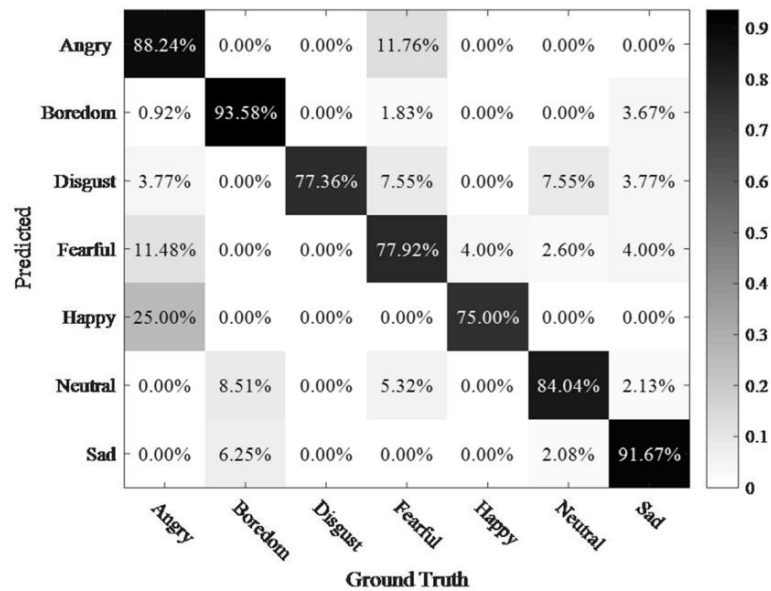
Basis data Te Berlin [37] adalah basis data ucapan emosional di Jerman, yang menyediakan 535 audio ucapan sehari-hari yang mengandung tujuh emosi: (1) "marah", (2) "kebosanan", (3) "jijik", (4) "takut", (5) "senang", (6) "netral", dan (7) "sedih". Data direkam pada laju pengambilan sampel 48-kHz dan kemudian diturunkan sampelnya menjadi 16-kHz. Contoh database Berlin disajikan pada Gambar 7. Gambar tersebut mewakili ucapan yang sama yang diucapkan oleh seorang Jerman dalam tiga emosi berbeda. Spektrum di atas mewakili ketakutan, spektrum tengah menunjukkan kebahagiaan, dan spektrum terakhir menunjukkan kemarahan.

Untuk mengevaluasi kinerja pendekatan yang diusulkan, berbagai metrik dihitung menggunakan (1)  $\bar{y}$  (5), di mana akurasi (ACC) mewakili keakuratan model [38,





**Gambar 8** Kinerja SEC-MLP tanpa augmentasi data menggunakan fitur paralinguistik. ACC = 62,95%, TPR = 39,72%, TNR = 79,26%, PPV = 57,36%, NPV = 65,18%. Hasil ACC digunakan untuk membuat matriks konfusi

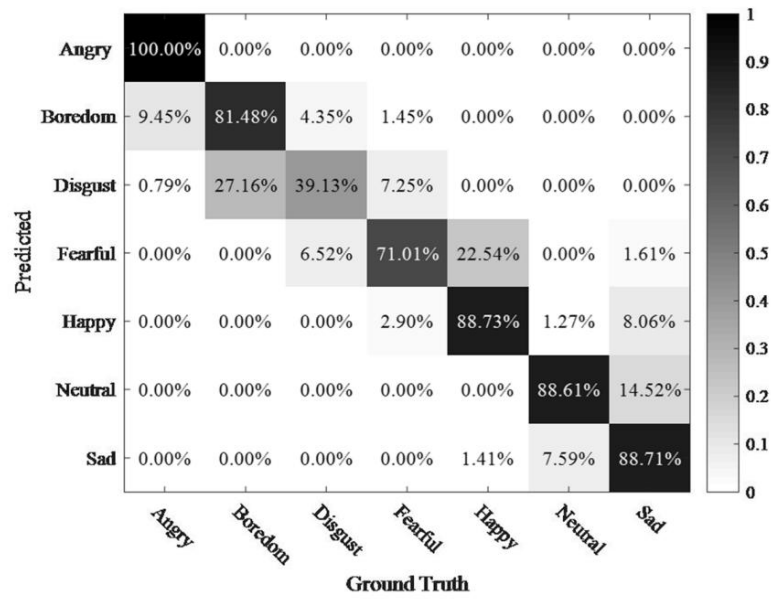


**Gambar 9** Kinerja SEC-MLP dengan augmentasi data menggunakan fitur paralinguistik. ACC = 83,97%, TPR = 69,07%, TNR = 91,28%, PPV = 79,52%, NPV = 85,75%. Hasil ACC digunakan untuk membuat matriks konfusi

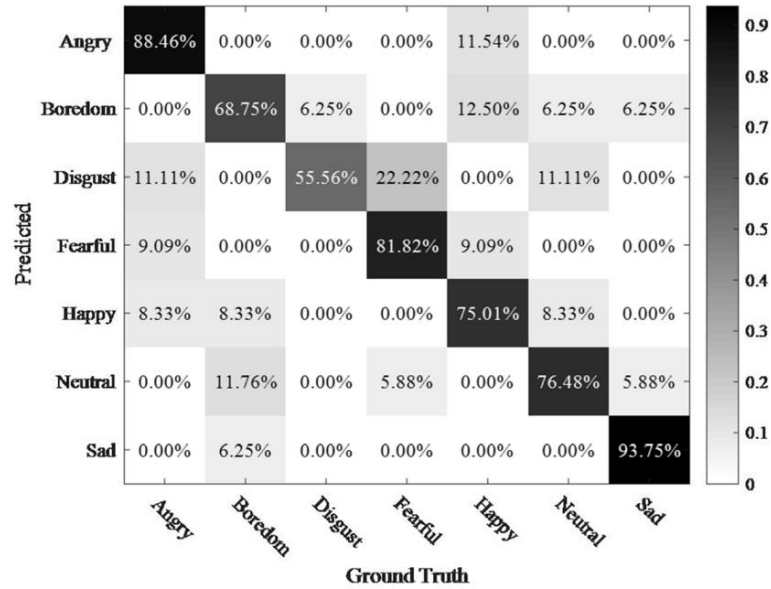
39], tingkat positif sebenarnya (TPR) mengukur seberapa baik model mendeteksi peristiwa positif, dan tingkat negatif sebenarnya (TNR) menunjukkan seberapa tepat penugasan ke kelas positif. Nilai prediksi positif (PPV) menunjukkan seberapa tepat model dalam menetapkan peristiwa positif ke kelas positif, dan nilai prediksi negatif (NPV) mengukur seberapa tepat model tersebut menentukan kelas positif.

akurat model dalam mendeteksi peristiwa negatif. TP, TN, FP, dan FN masing-masing merupakan label prediksi positif benar, negatif benar, positif palsu, dan negatif palsu.

$$ACC = \frac{TP + TN}{P + N} \quad (1)$$



**Gambar 10** Kinerja SEC-SVM menggunakan fitur paralinguistik. ACC = 83,74%, TPR = 90,71%, TNR = 96,13%, PPV = 99,03%, NPV = 96,11%. Hasil ACC digunakan untuk membuat matriks konfusi



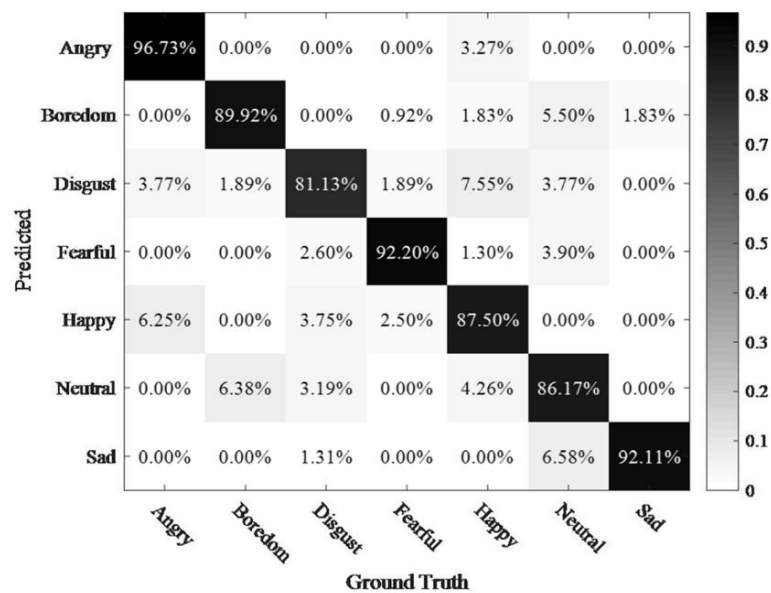
**Gambar 11** Kinerja SEC-MLP tanpa augmentasi data menggunakan fitur spektral. ACC = 77,12%, TPR = 57,57%, TNR = 88,90%, PPV = 75,75%, NPV = 77,67%. Hasil ACC digunakan untuk membuat matriks konfusi

$$TPR = \frac{dII}{P} = \frac{dII}{TP + FN} \quad (2)$$

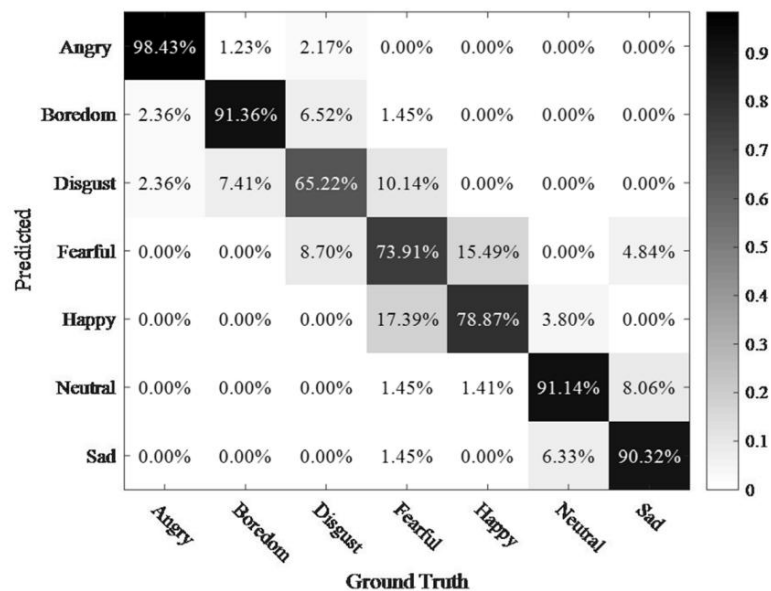
$$PPV = \frac{dII}{TP + FP} \quad (4)$$

$$TNR = \frac{TN}{N} = \frac{TN}{FP + TN} \quad (3)$$

$$NPV = \frac{TN}{TN + FN} \quad (5)$$



**Gambar 12** Kinerja SEC-MLP dengan augmentasi data menggunakan fitur spektral. ACC = 89,39%, TPR = 78,38%, TNR = 94,5%, PPV = 86,84%, NPV = 90,42%. Hasil ACC digunakan untuk membuat matriks konfusi



**Gambar 13** Kinerja SEC-SVM menggunakan fitur spektral. ACC = 86,73%, TPR = 95,42%, TNR = 99,41%, PPV = 98,43%, NPV = 98,26%. Hasil ACC digunakan untuk membuat matriks konfusi

### 3.1 Kinerja penggunaan pendekatan yang diusulkan ciri paralinguistik

Kami menunjukkan kinerja pendekatan klasifikasi emosi ucapan yang diusulkan menggunakan fitur paralinguistik di bagian ini, seperti yang ditunjukkan pada Gambar. 8, 9, dan 10, dimana hasil ACC dihitung dan digunakan untuk membuat matriks konfusi. Gambar 8 dan 9 menunjukkan hasilnya

SEC-MLP tanpa dan dengan augmentasi data masing-masing. Hasilnya menunjukkan bahwa kinerja pendekatan ini dapat meningkat pesat dengan operasi augmentasi data. Selain hasil ACC, kami juga menghitung hasil TPR, TNR, PPV, dan NPV dari setiap klasifikasi emosi ucapan, dan hasil rata-rata dihitung tanpa augmentasi data.

Tabel 1 Perbandingan skema yang diusulkan dengan pekerjaan yang ada dalam kondisi bersih

		ACC	TPR	TNR	PPV	NPV
Metode [39]		78,07%	88,19%	96,60%	84,35%	97,64%
SEC-MLP menggunakan fitur paralinguistik	Tanpa augmentasi data	62,95%	39,72%	79,26%	57,36%	65,18%
	Dengan augmentasi data	83,97%	69,07%	91,28%	79,52%	85,75%
SEC-SVM menggunakan fitur paralinguistik		83,74%	90,71%	96,13%	99,03%	96,11%
SEC-MLP menggunakan fitur spektral	Tanpa augmentasi data	77,12%	57,57%	88,90%	75,75%	77,67%
	Dengan augmentasi data	89,39%	78,38%	94,50%	86,84%	90,42%
SEC-SVM menggunakan fitur spektral		86,73%	95,42%	99,41%	98,43%	98,26%

Data yang dicetak tebal menunjukkan hasil terbaik

Tabel 2 Perbandingan skema yang diusulkan dengan pekerjaan yang ada dalam kondisi bising

SNR	metode	ACC	TPR	TNR	PPV	NPV
5dB	Metode [39]	82,55%	91,68%	93,72%	82,60%	97,36%
	SEC-SVM menggunakan fitur paralinguistik	66,17%	97,87%	88,22%	72,44%	99,24%
	SEC-SVM menggunakan fitur spektral	57,38%	69,40%	100,00%	100,00%	76,27%
10dB	Metode [39]	84,52%	93,29%	94,52%	84,62%	97,89%
	SEC-SVM menggunakan fitur paralinguistik	67,48%	99,04%	91,49%	81,10%	99,37%
	SEC-SVM menggunakan fitur spektral	71,59%	82,47%	100,00%	100,00%	90,46%
15dB	Metode [39]	84,68%	92,58%	94,62%	84,94%	97,68%
	SEC-SVM menggunakan fitur paralinguistik	74,39%	94,40%	96,89%	92,91%	97,56%
	SEC-SVM menggunakan fitur spektral	83,55%	92,70%	99,00%	97,27%	96,97%

Data yang dicetak tebal menunjukkan hasil terbaik

dan dengan augmentasi data masing-masing. Bila tanpa augmentasi data, rata-rata ACC, TPR, TNR, PPV, dan NPV masing-masing adalah 62,95%, 39,72%, 79,26%, 57,36%, dan 65,18%, sedangkan bila diterapkan augmentasi data, rata-rata ACC, TPR, TNR, PPV, dan NPV masing-masing sebesar 83,97%, 69,07%, 91,28%, 79,52%, dan 85,75%. masing-masing. Hasil keseluruhan menunjukkan bahwa augmentasi data sangat meningkatkan hasil klasifikasi.

Gambar 10 menunjukkan kinerja SEC-SVM yang diusulkan, di mana data gabungan dari database Berlin dipecah menjadi sepuluh subset dengan algoritma validasi silang sepuluh kali lipat, dan sembilan dari sepuluh ditransmisikan sebagai masukan SVM untuk model kereta. Label prediksi dihasilkan oleh prediksi SVM setelah memasukkan data pengujian, model, dan opsi prediksi. Rata-rata ACC, TPR, TNR, PPV, dan NPV dihitung masing-masing sebesar 83,74%, 90,71%, 96,13%, 99,03%, dan 96,11%. Dapat dengan mudah dilihat bahwa SEC-SVM berkinerja lebih baik daripada SEC-MLP dalam kondisi yang sama.

3.2 Kinerja pendekatan yang diusulkan menggunakan spektral fitur

Mirip dengan Bagian 3.1, pada bagian ini, kami menunjukkan kinerja pendekatan klasifikasi emosi ucapan yang diusulkan menggunakan fitur spektral, dan hasilnya adalah

ditunjukkan pada Gambar. 11, 12 dan 13, dimana hasil ACC digunakan untuk membuat matriks konfusi. Gambar 11 dan 12 masing-masing menunjukkan hasil SEC-MLP tanpa dan dengan augmentasi data. Kami juga menghitung hasil TPR, TNR, PPV, dan NPV dari setiap klasifikasi emosi ucapan, dan hasil rata-rata dihitung masing-masing tanpa augmentasi data dan dengan augmentasi data. Bila tanpa augmentasi data, rata-rata ACC, TPR, TNR, PPV, dan NPV masing-masing adalah 77,12%, 57,57%, 88,90%, 75,75%, dan 77,67%, sedangkan bila diterapkan augmentasi data, rata-rata ACC, TPR, TNR, PPV, dan NPV masing-masing sebesar 89,39%, 78,38%, 94,5%, 86,84%, dan 90,42%. masing-masing. Hasilnya menunjukkan bahwa kinerja pendekatan ini dapat meningkat pesat dengan operasi augmentasi data.

Gambar 13 menunjukkan kinerja SEC-SVM yang diusulkan, di mana data gabungan dari database Berlin dipecah menjadi sepuluh subset dengan algoritma validasi silang sepuluh kali lipat, dan sembilan dari sepuluh ditransmisikan sebagai input SVM untuk model kereta api. Rata-rata ACC, TPR, TNR, PPV, dan NPV dihitung masing-masing sebesar 86,73%, 95,42%, 99,41%, 98,43%, dan 98,26%. Dapat dengan mudah dilihat bahwa SEC-SVM berkinerja lebih baik daripada SEC-MLP dalam kondisi yang sama.

3.3 Perbandingan skema yang diusulkan dengan skema yang ada metode

Untuk menguji kinerja skema yang diusulkan dalam lingkungan ucapan nyata, pada bagian ini, kami membandingkan kinerja skema yang diusulkan dengan metode yang ada [39], masing-masing ketika dalam kondisi bersih dan kondisi bising, dan menunjukkan hasilnya pada Tabel 1 dan 2. Pada Tabel 2, white noise Gaussian dengan rasio signal-to-noise yang berbeda (5, 10, 15 dB) digunakan untuk mendistorsi sampel ucapan pengujian sebelum ekstraksi fitur. ACC, TPR, TNR, PPV, dan NPV dihitung untuk menunjukkan kinerja skema yang diusulkan.

Pada Tabel 1, kinerja SEC-MLP dan SEC-SVM yang diusulkan dihitung, masing-masing menggunakan fitur paralinguistik dan fitur spektral. Metrik ACC, TPR, TNR, PPV, NPV dihitung, dan hasil terbaik dari setiap metrik ditandai dengan huruf tebal untuk kejelasan. Hasilnya menunjukkan bahwa SEC-MLP yang menggunakan fitur paralinguistik mencapai hasil yang jauh lebih baik dengan augmentasi data, dalam hal semua ACC, TPR, TNR, PPV, dan NPV. Di sisi lain, SEC-SVM yang menggunakan fitur spektral memberikan hasil yang lebih baik dalam hal TPR, TNR, dan NPV. Pada Tabel 2, kami mensimulasikan lingkungan nyata dengan menambahkan white noise Gaussian dengan rasio signal-to-noise yang berbeda (5, 10, 15 dB) dan menghitung metrik ACC, TPR, TNR, PPV, dan NPV.

4 Kesimpulan dan pekerjaan masa depan

Dalam makalah ini, kami mengusulkan pendekatan untuk klasifikasi emosi bicara dan kami mengevaluasi pendekatan tersebut dalam berbagai kasus. Fitur paralinguistik tradisional berdasarkan rangkaian tantangan paralinguistik INTERSPEECH 2013 diekstraksi menggunakan toolkit OpenSMILE. Fitur spektral diekstraksi oleh MFCC yang terkenal. Pengklasifikasi SVM dan MLP digunakan masing-masing untuk klasifikasi. Eksperimen telah dilakukan pada database Berlin untuk mengevaluasi kinerja pendekatan yang diusulkan. ACC, TPR, TNR, PPV, dan NPV masing-masing dihitung untuk mengukur pendekatan yang diusulkan. Hasil percobaan menunjukkan bahwa pendekatan yang diusulkan mencapai kinerja yang baik dalam kondisi yang berbeda dan berkinerja lebih baik dibandingkan pekerjaan terkait dalam hal berbagai metrik evaluasi. Pekerjaan kami di masa depan akan fokus pada penggabungan berbagai fitur untuk kinerja klasifikasi yang lebih baik dan penerapan teknik pembelajaran mendalam untuk klasifikasi.

Singkatan

MFCC	Koefisien cepstral frekuensi mel
SVM	Mendukung mesin vektor
LSPC	Catat koefisien daya frekuensi
ELM	Mesin pembelajaran ekstrim
MLP	Perceptron multi-lapis
LLD	Deskriptor tingkat rendah

Klasifikasi emosi ucapan SEC-MLP menggunakan pengklasifikasi MLP

Klasifikasi emosi ucapan SEC-SVM menggunakan SVM

LLCF	Pemilihan fitur pengelompokan
UDFS	Pemilihan fitur diskriminatif tanpa pengawasan
ACC	Ketepatan
TPR	Tingkat positif yang sebenarnya
TNR	Tingkat negatif sebenarnya
PPV	Nilai prediksi positif
NPV	Nilai prediksi negatif

Ucapan Terima Kasih

Tak dapat diterapkan.

Kontribusi penulis

Konseptualisasi: Tong Liu dan Xiaochen Yuan; kurasi data: Tong Liu; analisis formal: Tong Liu; perolehan pendanaan: Xiaochen Yuan; metodologi: Tong Liu dan Xiaochen Yuan; administrasi proyek: Xiaochen Yuan; pengawasan: Xiaochen Yuan; validasi: Tong Liu; visualisasi: Tong Liu; tulisany draf asli: Tong Liu; menulisyresensi dan penyuntingan: Xiaochen Yuan. Para penulis membaca dan menyetujui naskah akhir.

Pendanaan

Penelitian ini didanai oleh proyek Penelitian Universitas Politeknik Macao (Proyek No. RP/ESCA-03/2021, RP/FCA-12/2022).

Ketersediaan data dan bahan

Dataset yang kami gunakan dalam pekerjaan ini adalah dataset Berlin, yang juga dikenal sebagai database EMODB, dan tersedia di: <https://www.kaggle.com/datasets/piyuhagni5/berlin-database-of-emosional-speech-emodb>.

Deklarasi

Persetujuan etika dan persetujuan untuk berpartisipasi

Tak dapat diterapkan.

Persetujuan untuk publikasi

Tak dapat diterapkan.

Kepentingan yang bersaing

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan bersaing.

Diterima: 27 Februari 2023 Diterima: 29 April 2023  
Published online: 15 May 2023

Referensi

1. X. Cao, M. Jia, J. Ru, Tw Pai, Pengenalan emosi ucapan lintas korpus menggunakan pembelajaran subruang dan adaptasi domain. EURASIP J. Proses Musik Pidato Audio. 2022(1), 32 (2022)
2. K. Wang, N. An, BN Li, Y. Zhang, L. Li, Pengenalan emosi ucapan menggunakan parameter fourier. IEEE Trans. Mempengaruhi. Hitung. 6(1), 69–75 (2015)
3. D. Tang, P. Kuppens, L. Geurts, T. van Waterschoot, Pengenalan emosi ucapan ujung ke ujung menggunakan jaringan saraf konvolusi dilatasi susunan konteks baru. EURASIP J. Proses Musik Pidato Audio. 2021(1), 18 (2021)
4. L. Sun, S. Fu, F. Wang, Model svm pohon keputusan dengan pemilihan fitur lebih lanjut untuk pengenalan emosi ucapan. EURASIP J. Proses Musik Pidato Audio. 2019(1), 1–14 (2019)
5. P. Ekman, Argumen tentang emosi dasar. Pengetahuan. emosi. 6(3–4), 169–200 (1992)
6. JA Russell, Model pengaruh sirkumplex. J.Pers. sosial. Psikologi. 39(6), 1161 (1980)
7. A. Cabri, F. Masulli, Z. Mnasri, S. Rovetta dkk., Pengenalan emosi dari pidato: pendekatan pembelajaran tanpa pengawasan. Int. J.Komputasi. Intel. sistem. 14(1), 23 (2020)
8. KR Scherer et al., Tentang sifat dan fungsi emosi: pendekatan proses komponen. Mendekati Emosi. 2293(317), 31 (1984)
9. Rao, KS, Koolagudi, SG Pengenalan emosi yang kuat menggunakan spektral dan fitur prosodik. Dalam: Sains & Media Bisnis Springer, Springer, NewYork (2013)



10. Y. Xie, R. Liang, Z. Liang, C. Huang, C. Zou, B. Schuller, Klasifikasi emosi ucapan menggunakan lsm berbasis perhatian. *IEEE/ACM Trans. Audio Pidato Lang. Proses.* 27(11), 1675–1685 (2019)
11. Y. Xu, W. Wang, H. Cui, M. Xu, M. Li, Pengenalan atribut nyanyian paralinguistik menggunakan pembelajaran mesin yang diawasi untuk mendeskripsikan suara nyanyian solo tenor klasik dalam pedagogi vokal. *EURASIP J. Proses Musik Pidato Audio.* 2022(1), 1–16 (2022)
12. E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, A. Stolcke, Pemodelan rangkaian fitur pro-sodik untuk pengenalan pembicara. *Pidato Komuni.* 46(3–4), 455–472 (2005)
13. SR Kshirsagar, TH Falk, Fitur spektrum modulasi yang sadar kualitas untuk pengenalan emosi ucapan yang kuat. *IEEE Trans. Mempengaruhi. Hitung.* 13(4), 1892–1905 (2022)
14. M. Geravanchizadeh, E. Forouhandeh, M. Bashirpour, Kompensasi fitur tion berdasarkan normalisasi panjang saluran vokal untuk peningkatan pengenalan ucapan yang dipengaruhi emosi. *EURASIP J. Proses Musik Pidato Audio.* 2021, 1–19 (2021)
15. JL Jacobson, DC Boersma, RB Fields, KL Olson, Fitur paralinguistik ucapan orang dewasa kepada bayi dan anak kecil. *Pengembang Anak.* 54(2), 436–442 (1983)
16. SM Tsai, pada Konferensi Internasional Pertama tentang Teknologi Oranye (ICOT) tahun 2013, Algoritma zero-watermarking yang kuat untuk audio berdasarkan LPCC (IEEE, 2013), hal. 63–66
17. C. Ittichaichareon, S. Sukrisi, T. Yingthawornusuk, dalam konferensi Internasional tentang grafik komputer, simulasi dan pemodelan (ICGSM'2012), vol. 9, Pengenalan ucapan menggunakan mfcc, Pattaya, Thailand (2012)
18. TL Nwe, SW Foo, LC De Silva, Pengenalan emosi ucapan menggunakan model markov tersembunyi. *Pidato Komuni.* 41(4), 603–623 (2003)
19. F. Albu, D. Hagiescu, L. Vladutu, MA Puica, dalam Prosiding EDULEARN15, Pendekatan jaringan saraf untuk pengenalan emosi anak-anak dalam aplikasi pembelajaran cerdas (IATED, 2015), hlm. 3229–3239
20. GB Huang, QY Zhu, CK Siew, Mesin pembelajaran ekstrim: teori dan aplikasi. *Komputasi saraf* 70(1–3), 489–501 (2006)
21. LE Peterson, K-tetangga terdekat. *Cendekiawanpedia* 4(2), 1883 (2009)
22. B. Schuller, G. Rigoll, M. Lang, pada konferensi internasional IEEE tahun 2004 tentang akustik, ucapan, dan pemrosesan sinyal, vol. 1, Pengenalan emosi ucapan yang menggabungkan fitur akustik dan informasi linguistik dalam arsitektur jaringan kepercayaan mesin vektor dukungan hibrid (IEEE, 2004), hal. I–577
23. K. Han, D. Yu, I. Tashev, dalam INTERSPEECH 2014, Pengenalan emosi ucapan menggunakan jaringan saraf dalam dan mesin pembelajaran ekstrim, ISCA, Singapura (2014)
24. B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, S. Naray-anan, di Proc. INTERSPEECH 2010, Tantangan paralinguistik INTERSPEECH 2010, ISCA, Makuhari, Jepang, (2010), hlm. 2794–2797
25. F. Eyben, M. Wöllmer, B. Schuller, dalam Prosiding konferensi internasional ACM ke-18 tentang Multimedia, Opensmile: ekstraktor fitur audio sumber terbuka serbaguna dan cepat dari munich, ACM, New York, Amerika Serikat (2010), hal.1459–1462
26. J. Joy, A. Kannan, S. Ram, S. Rama, Pengenalan emosi ucapan menggunakan jaringan saraf dan pengklasifikasi MLP, *International Journal of Engineering Science and Computing*, Pearl Media Publications PVT LTD, 10(4), hlm.25170 –25172 (2020)
27. B. McFee, C. Rafel, D. Liang, DP Ellis, M. McVicar, E. Battenberg, O. Nieto, dalam Proceedings of the 14th python in science conference, vol. 8, librosa: analisis sinyal audio dan musik dengan python (Citeseer, 2015), hlm.18–25
28. F. Albu, A. Mateescu, N. Dumitriu, dalam Konferensi Internasional tentang Mikro-elektronik dan Ilmu Komputer, Pemilihan arsitektur untuk jaringan feedforward multilayer (Citeseer, 1997), hlm. 131–134
29. C. Xiang, SQ Ding, TH Lee, Interpretasi geometris dan pemilihan arsitektur MLP. *IEEE Trans. Jaringan Syaraf.* 16(1), 84–96 (2005)
30. T. Andersen, T. Martinez, dalam IJCNN'99. Konferensi Gabungan Internasional tentang Jaringan Syaraf Tiruan. Prosiding (Kat. No. 99CH36339), vol. 3, Validasi silang dan pemilihan arsitektur MLP (IEEE, 1999), hlm.1614–1619
31. G. Rofo, Pustaka pemilihan fitur (kotak alat matlab). *arXiv preprint arXiv:1607.01327* (2016)
32. S. Russell, P. Norvig, Kecerdasan buatan: pendekatan modern, Prentice Hall, London, Inggris (2003)
33. H. Peng, F. Long, C. Ding, Pemilihan fitur berdasarkan kriteria informasi timbal balik max-dependency, max-relevance, dan min-redundancy. *IEEE Trans. Pola Dubur. Mach. Intel.* 27(8), 1226–1238 (2005)
34. H. Liu, H. Motoda, dalam Chapman & Hall/CRC, Metode komputasi fitur seleksi masa depan (seri penambangan data dan penemuan pengetahuan Chapman & Hall/CRC), Chapman dan Hall/CRC, Florida, Amerika Serikat (2007)
35. H. Zeng, Ym Cheung, Pemilihan fitur dan pembelajaran kernel untuk pembelajaran lokal-pengelompokan berbasis ing. *IEEE Trans. Pola Dubur. Mach. Intel.* 33(8), 1532–1547 (2010)
36. Y. Yang, HT Shen, Z. Ma, Z. Huang, X. Zhou, dalam internasional ke dua puluh dua konferensi gabungan tentang kecerdasan buatan, L2, pemilihan fitur diskriminatif yang diatur dengan 1 norma untuk tanpa pengawasan, AAAI Press, Washington, Amerika Serikat (2011)
37. F. Burkhardt, A. Paeschke, M. Rolfes, WF Sendlimeier, B. Weiss, dkk., dalam INTERSPEECH, jilid. 5, Database pidato emosional Jerman. ISCA, Lisbon, Portugal (2005), hlm.1517–1520
38. Y. Fu, X. Yuan, pada Konferensi Internasional Komputasi IEEE ke-23 tahun 2020 Sains dan Teknik (CSE), Ekstraksi fitur komposit untuk pengenalan emosi ucapan (IEEE, 2020), hal. 72–77
39. WA Jassim, R. Paramesran, N. Harte, Klasifikasi emosi ucapan menggunakan gabungan neurogram dan fitur tantangan paralinguistik INTERSPEECH 2010. *Proses Sinyal IET.* 11(5), 587–595 (2017)

## Catatan Penerbit

Springer Nature tetap netral sehubungan dengan klaim yurisdiksi dalam peta yang diterbitkan dan afiliasi kelembagaan.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)