

EKSTRAKSI FITUR PENGENALAN EMOSI BERDASARKAN UCAPAN MENGGUNAKAN *LINEAR PREDICTOR CEPTRAL COEFFECIENT* DAN *MEL FREQUENCY CEPSTRUM COEFFICIENTS*

Siti Helmiyah^{1,*}, Imam Riadi², Rusydi Umar³, Abdullah Hanif⁴

^{1,3,4}Program Studi Teknik Informatika, Universitas Ahmad Dahlan Yogyakarta

²Program Studi Sistem Informasi, Universitas Ahmad Dahlan Yogyakarta

e-mail: ^{1,*}siti1708048022@webmail.uad.ac.id ²imam_riadi@is.uad.ac.id

^{*}) correspondence email

Abstrak

Ucapan suara memiliki informasi penting yang dapat diterima oleh otak melalui gelombang suara. Otak menerima gelombang suara melalui alat pendengaran dan menghasilkan suatu informasi berupa pesan, bahasa, dan emosi. Pengenalan emosi wicara merupakan teknologi yang dirancang untuk mengidentifikasi keadaan emosi seseorang dari sinyal ucapannya. Hal tersebut menarik untuk diteliti, karena berkaitan dengan teknologi zaman sekarang yaitu pada penggunaan *smartphone* di berbagai macam aktivitas sehari-hari. Penelitian ini membandingkan ekstraksi fitur Metode LPC dan Metode MFCC. Kedua metode ekstraksi tersebut diklasifikasi menggunakan Metode Jaringan Syaraf Tiruan (MLP) untuk pengenalan emosi. Masing-masing metode menggunakan data emosi marah, bosan, bahagia, netral, dan sedih. Data dibagi menjadi dua, yaitu data testing dan data data training dengan perbandingan 80:20. Arsitektur jaringan yang digunakan adalah tiga lapisan yaitu lapisan input, lapisan tersembunyi, dan lapisan output. Parameter MLP yang digunakan learning rate = 0.0001, epsilon = 1e-08, epoch = 500, dan Cross Validation = 5. Hasil akurasi pengenalan emosi dengan ekstraksi fitur LPC sebesar adalah 28%. Sedangkan hasil akurasi dengan ekstraksi fitur MFCC sebesar 61,33%. Hasil akurasi ini bisa ditingkatkan dengan menambahkan data yang lebih banyak lagi, terutama untuk data testing. Perlunya pengujian pada nilai parameter jaringan MLP, yaitu dengan mengubah nilai-nilai parameter, karena dapat mempengaruhi tingkat akurasi pengenalan. Selain itu penentuan ekstraksi fitur dan klasifikasi metode yang lain juga dapat digunakan untuk mencari nilai akurasi pengenalan emosi yang lebih baik lagi.

Kata Kunci: Emosi, Ucapan, LPC, MFCC, MLP

PENDAHULUAN

Karakteristik manusia adalah berinteraksi dengan sesama. Sudah ribuan tahun manusia berinteraksi satu sama lain dengan berbicara. Akhir-akhir ini pengembangan interaksi manusia dan komputer melalui ucapan suara, pada bidang akademis dan industri menjadi salah satu topik dan inovasi yang menarik [1]. Ucapan suara adalah salah satu cara berkomunikasi dengan sesama manusia. Ucapan suara memiliki informasi penting yang dapat diterima oleh otak melalui gelombang suara. Otak menerima gelombang suara melalui alat pendengaran dan menghasilkan suatu informasi berupa pesan, bahasa, dan emosi [2].

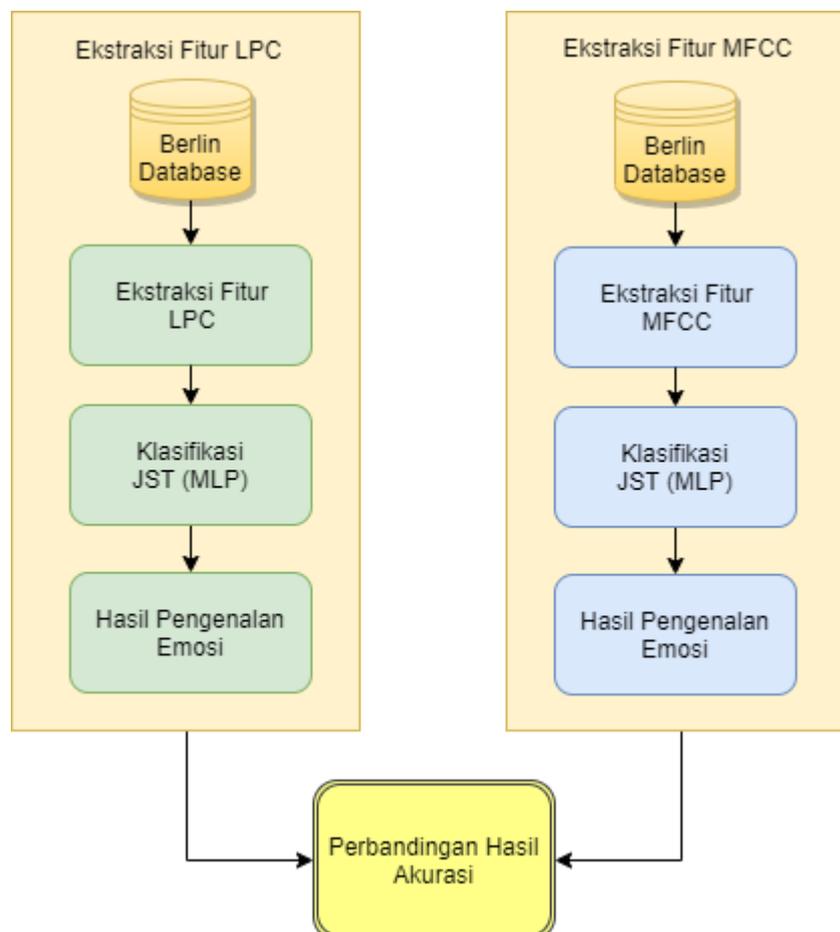
Emosi adalah suatu keadaan yang kompleks pada makhluk hidup [3]. Emosi juga dapat dipengaruhi oleh kondisi psikologis dan keadaan suasana hati seseorang. Penelitian [4] berpendapat ada lima dasar emosi yaitu marah, bahagia, sedih, jijik, dan takut. Pengenalan emosi wicara adalah teknologi yang dirancang untuk mengidentifikasi keadaan emosi seseorang dari sinyal ucapannya [5]. Hal tersebut menarik untuk diteliti, karena berkaitan dengan teknologi zaman sekarang yaitu pada penggunaan *smartphone* di berbagai macam aktivitas sehari-hari.

Ekstraksi fitur adalah tahapan paling penting dalam pengenalan ucapan emosi. Ekstraksi fitur yang baik adalah dapat membedakan pola fitur antara satu kelas emosi dengan kelas emosi lain. Menurut [6] pengenalan suara ucapan emosi dibagi menjadi tiga yaitu fitur prosodis, fitur spektral, dan fitur kualitas suara. Fitur yang sering digunakan dalam penelitian antara lain MSF (*Modulation Spectral Feature*), MFCC (*Mel Frequency Cepstrum Coefficients*), ZCPA (*Zero Crossings with Peak Amplitudes*), dan LPCC (*Linear Predictor Ceptral Coefficient*) [7].

Penelitian ini dilakukan untuk membandingkan hasil akurasi dari pengenalan emosi, pada dua metode ekstraksi fitur yang berbeda. Metode ekstraksi fitur yang digunakan yaitu Metode *Linear Predictive Coding* (LPC) dan Metode *Mel Frequency Cepstrum Coefficients* (MFCC). Setelah proses dari kedua metode ekstraksi fitur tersebut dilakukan, masing-masing hasil dari fitur tersebut diklasifikasikan menggunakan metode Jaringan Syaraf Tiruan yaitu *Multilayer Perceptron* (MLP).

METODE PENELITIAN

Perancangan sistem untuk perbandingan metode ekstraksi fitur untuk pengenalan emosi ucapan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Pengenalan Emosi dengan Membandingkan Metode Ekstraksi Fitur

Pada Gambar 1 Tahapan utama yang paling penting dalam pengenalan emosi ucapan adalah proses ekstraksi fitur dan klasifikasi. Dua tahapan ini sangat berpengaruh dalam menentukan hasil pengenalan emosi. Tahapan awal adalah input

data emosi ucapan suara menggunakan *Berlin Database of Emotional Speech Berlin Emo-DB*. Data suara di ekstraksi menggunakan fitur spektral. Pada penelitian ini ekstraksi fitur spektral yang digunakan adalah ekstraksi fitur metode LPC dan metode MFCC. Hasil ekstraksi fitur tersebut masing-masing diklasifikasi menggunakan metode Jaringan Syaraf Tiruan (*MultiLayer Perceptron*).

A. Database Emosi

Data ucapan suara emosi yang digunakan pada penelitian ini adalah Berlin Database of Emotional Speech Berlin Emo-DB. Berlin Emo-DB adalah salah satu proyek penelitian DFG SE462/3-1 pada tahun 1997. Prof. Dr. W. Sendlmeier dari *Technical University of Berlin* adalah direktur pada proyek tersebut. Rekaman untuk pembuatan database dikerjakan di ruang anechoic di *Technical University Berlin, Department of Technical Acoustics*. Proses pemilihan aktor dilakukan melalui proses seleksi dan menghasilkan 10 orang aktor, lima laki-laki dan lima perempuan. Umur 10 aktor tersebut berkisar antara 21 sampai 35 tahun. Data kode, jenis kelamin, dan umur aktor dapat dilihat pada Tabel 1.

Tabel 1. Data Aktor Berlin Database

| Kode | Jenis Kelamin, Umur |
|-------------|----------------------------|
| 03 | male, 31 years old |
| 08 | female, 34 years |
| 09 | female, 21 years |
| 10 | male, 32 years |
| 11 | male, 26 years |
| 12 | male, 30 years |
| 13 | female, 32 years |
| 14 | female, 35 years |
| 15 | male, 25 years |
| 16 | female, 31 years |

Pada saat pembuatan *Berlin Database* dilakukan beberapa pertimbangan untuk pemilihan kalimat yang diucapkan. Kalimat yang diucapkan adalah kalimat yang digunakan sehari-hari. Pemilihan kalimat ini dilakukan, agar aktor dapat mengucapkan kalimat tersebut tanpa dibuat-buat. Pada Berlin Database, kalimat yang diucapkan terbagi menjadi dua, yaitu kalimat pendek dengan kode *a* dan kalimat panjang dengan kode *b*. Data kalimat tersebut diucapkan dengan emosi marah, bosan, bahagia, netral, sedih, jijik, dan takut. Data kalimat yang digunakan pada penelitian ini dapat dilihat pada

Tabel 2. Informasi Kalimat Berlin Database

| Kode | Kalimat Berlin Database | Terjemahan Inggris |
|-------------|--|---|
| a01 | Der Lappen liegt auf dem Eisschrank. | The tablecloth is lying on the frigde. |
| a02 | Das will sie am Mittwoch abgeben. | She will hand it in on Wednesday. |
| a04 | Heute abend könnte ich es ihm sagen. | Tonight I could tell him. |
| a05 | Das schwarze Stück Papier befindet sich da oben neben dem Holzstück. | The black sheet of paper is located up there besides the piece of timber. |

| | | |
|-----|--|--|
| a07 | In sieben Stunden wird es soweit sein. | In seven hours it will be. |
| b01 | Was sind denn das für Tüten, die da unter dem Tisch stehen? | What about the bags standing there under the table? |
| b02 | Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter. | They just carried it upstairs and now they are going down again. |
| b03 | An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht. | Currently at the weekends I always went home and saw Agnes. |

B. Ekstraksi Fitur

Tahapan paling penting dalam pengenalan ucapan emosi adalah penentuan ekstraksi fitur. Ekstraksi fitur yang baik adalah dapat membedakan pola fitur antara satu kelas emosi dengan kelas emosi lain. Ekstraksi fitur pada pengenalan suara ucapan emosi dapat dibagi menjadi tiga yaitu fitur prosodis, fitur spektral, dan fitur kualitas suara [6]. Fitur prosodis adalah unsur-unsur *phonetic* yang disuarakan. Fitur tersebut antara lain frekuensi, durasi, energi, pitch, formant, dan lain-lain. Fitur spektral adalah hasil dari interaksi antara energi elektromagnetik dengan suatu objek. Fitur spektral yang sering digunakan dalam penelitian yaitu *Linear Prediction Coefficient* (LPC), *Linear Prediction Cepstrum Coefficient* (LPCC), dan *Mel Frequency Cepstrum Coefficients* (MFCC) [8]. Fitur kualitas suara adalah hasil kualitas dari rekaman suara, seperti format frekuensi dan *bandwith*, *shimmer*, dan *jitter*. Fitur-fitur tersebut menghasilkan nilai dalam bentuk vektor yang digunakan pada proses klasifikasi. Penelitian ini membandingkan dua metode ekstraksi fitur yaitu *Linear Predictive Coding* (LPC) dan MFCC (*Mel Frequency Cepstrum Coefficients*).

Ekstraksi fitur LPC adalah salah satu metode dalam pengenalan ucapan. Ekstraksi fitur ini dapat diterapkan dalam pengenalan kata. Metode LPC mudah langsung diterapkan pada perangkat lunak maupun perangkat keras, karena perhitungan matematisnya relatif lebih singkat dari metode filter *bank* yang lain [9]. Selain itu, pemodelan LPC pada sinyal ucapan menghasilkan pendekatan vocal yang baik bagian *voiced* (bersuara), sedangkan untuk *unvoiced* (tidak ada suara), tidak begitu baik tapi masih dapat digunakan untuk pengenalan ucapan. Ekstraksi fitur ini mencari nilai koefisien LPC dari sinyal ucapan suara. Pada tahapan LPC biasanya orde (p) yang digunakan adalah 8 sampai 16 (Thiang, 2005). Orde ke-0 merupakan energi dari *frame* data suara. Pada tahap penggunaan orde menghasilkan koefisien $p+1$ [9].

Ekstraksi fitur MFCC adalah fitur yang sering digunakan dalam pengenalan suara. Tujuan dari metode MFCC adalah untuk meniru cara kerja telinga manusia dengan menerapkan analisis cepstral. Ekstraksi fitur MFCC dihitung berdasarkan *frame* ucapan. Biasanya, jumlah koefisien MFCC yang digunakan antara 9 sampai 13. Hal ini karena sebagian besar sinyal energi dipadatkan dalam beberapa koefisien pertama karena sifat dari transformasi cosinus [10].

C. Klasifikasi Emosi

Klasifikasi merupakan termasuk tahapan penting karena pada tahapan adalah penentuan hasil dari proses ekstraksi fitur yang telah dilakukan sebelumnya. Klasifikasi emosi pada penelitian ini menggunakan metode Jaringan Syaraf Tiruan: Multilayer Perceptron (MLP). Arsitektur jaringan yang dibuat terdiri tiga lapisan yaitu satu lapisan input, satu atau lebih lapisan tersembunyi, dan satu lapisan output. Klasifikasi dilakukan dua kali, yaitu pertama klasifikasi dengan menggunakan ekstraksi fitur LPC dan yang kedua klasifikasi dengan menggunakan ekstraksi fitur

MFCC. Hasil dari ekstraksi fitur dari metode LPC dan MFCC digunakan sebagai data input pada lapisan input.

Klasifikasi Jaringan Syaraf Tiruan (MLP) memiliki dua tahapan yaitu tahap *testing* dan tahap *training*. Tahap *testing* dilakukan untuk menentukan parameter-parameter yang terbaik, untuk digunakan pada tahap *training*. Penentuan parameter MLP seperti menentukan nilai *learning rate* (laju pembelajaran), jumlah iterasi (*epoch*), dan nilai batas *error (epsilon)*. Pada tahap *testing*, sebaiknya data yang digunakan banyak untuk menghasilkan pola yang baik. Tahap *training* adalah tahap untuk mengetahui hasil klasifikasi dari parameter-parameter yang telah ditentukan pada tahap *testing*. Pada penelitian ini data dibagi menjadi 80:20, dimana 80% data digunakan untuk *testing* dan 20% data digunakan untuk *training*. Parameter yang digunakan untuk klasifikasi emosi menggunakan MLP dapat dilihat pada Tabel 3.

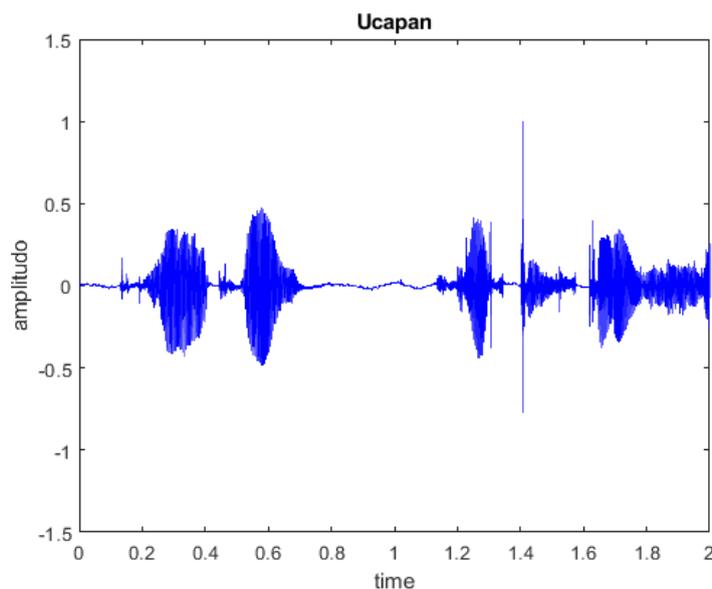
Tabel 3. Parameter Klasifikasi Emosi berdasarkan Jaringan Syaraf Tiruan

| Parameter | Nilai |
|--|----------|
| <i>Learning rate</i> (laju pembelajaran) | 0.0001 |
| Nilai batas <i>error (epsilon)</i> | 1e-08 |
| Maksimal Iterasi (<i>epoch</i>) | 500 |
| <i>CrossValidation</i> | 5 folds |
| Lapisan (<i>layer</i>) | 3 layers |
| Node Input LPC | 17 |
| Node Input MFCC | 13 |
| Node Lapisan Tersembunyi | 10 |
| Node Output | 5 |

HASIL DAN PEMBAHASAN

A. Data Emosi

Data ucapan suara emosi yang digunakan dalam penelitian ini menggunakan lima emosi yaitu emosi marah, bosan, bahagia, netral, dan sedih. Data yang digunakan sebanyak 250 data dari Berlin Database. Salah satu contoh gelombang sinyal dari data ucapan suara emosi Berlin Database dengan nama *file* 16a01Tb dapat dilihat pada Gambar 2.

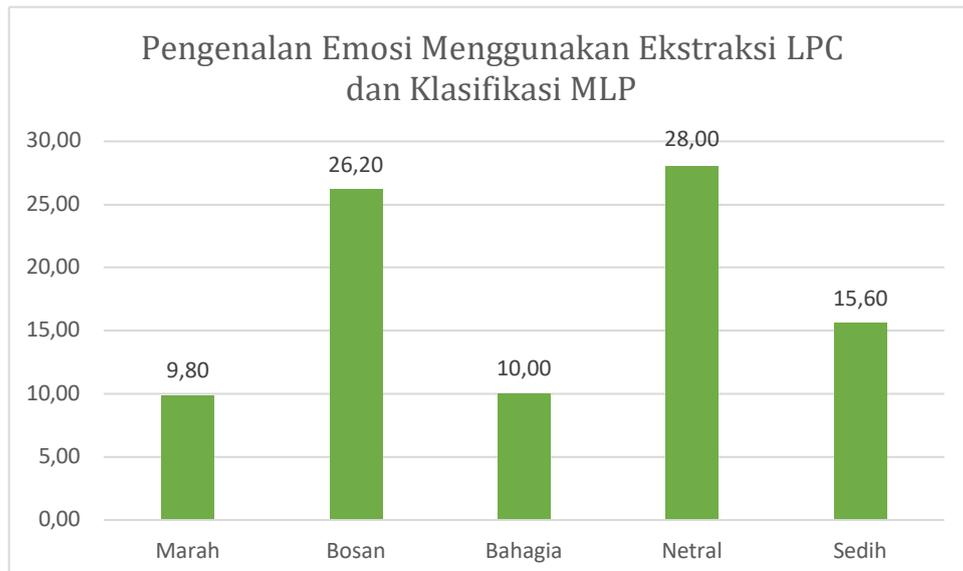


Gambar 2. Gelombang Sinyal Ucapan Suara Emosi

Pada Gambar 2, sumbu x adalah durasi (*time*) pengucapan kalimat dan sumbu y adalah tinggi rendah suara (*amplitude*). Keterangan *file 16a01Tb* yaitu *16* adalah kode aktor yang mengucapkan kalimat, *a01* adalah kode kalimat yang diucapkan, *T* adalah kode ucapan suara emosi sedih dan *b* merupakan kalimat yang diucapkan untuk kedua kali pada aktor tersebut.

B. Pengenalan Emosi Menggunakan Ekstraksi Fitur Metode LPC

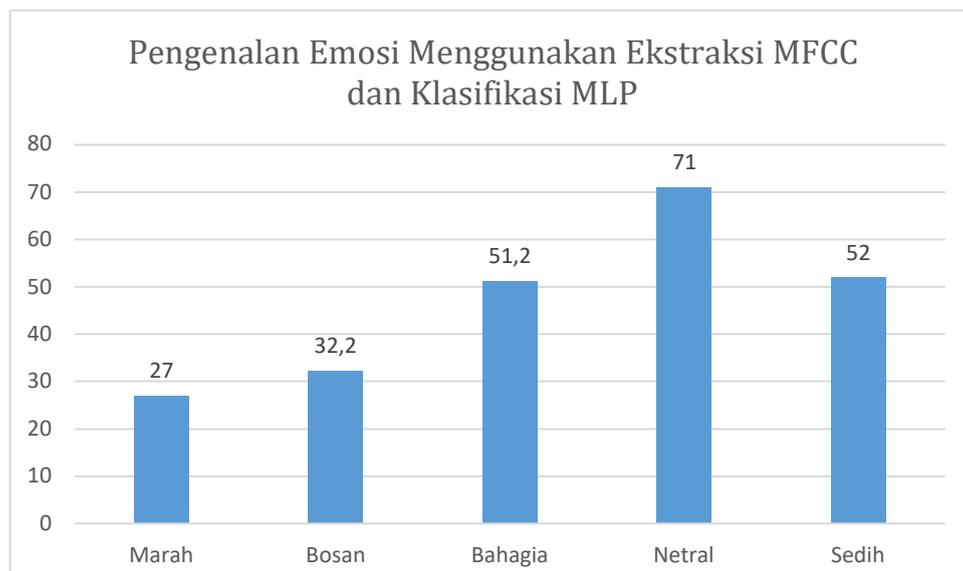
Rekaman ucapan suara emosi diekstraksi menggunakan LPC dengan menggunakan koefisien 16. Hasil dari ekstraksi ini menghasilkan vektor $[1 \times 17]$. Hasil ekstraksi fitur ini digunakan sebagai data input pada proses klasifikasi. Arsitektur jaringan menggunakan tiga lapisan dengan 17 node input pada lapisan input, 10 node pada lapisan tersembunyi, 5 node output yaitu pada lima kelas emosi pada lapisan output. Parameter MLP yang digunakan *learning rate* = 0.0001, *epsilon* = $1e-08$, *epoch* = 500 dan *CrossValidation* yang digunakan adalah 5 kelipatan. Hasil akurasi pengenalan emosi menggunakan ekstraksi LPC dan klasifikasi MLP dapat dilihat pada Gambar 3.



Gambar 3. Hasil Pengenalan Emosi Menggunakan Ekstraksi LPC dan Klasifikasi MLP

C. Pengenalan Emosi Menggunakan Ekstraksi Fitur Metode MFCC

Rekaman ucapan suara emosi diekstraksi menggunakan MFCC dengan koefisien 13. Hasil ekstraksi fitur ini menghasilkan vektor [1x13]. Hasil ekstraksi fitur tersebut digunakan sebagai data input pada proses klasifikasi. Arsitektur jaringan menggunakan tiga lapisan dengan 13 node input pada lapisan input, 10 node pada lapisan tersembunyi, 5 node output yaitu pada lima kelas emosi pada lapisan output. Parameter MLP yang digunakan *learning rate* = 0.0001, *epsilon* = 1e-08, *epoch* = 500 dan *CrossValidation* yang digunakan adalah 5 kelipatan. Hasil akurasi pengenalan emosi menggunakan ekstraksi MFCC dan klasifikasi MLP dapat dilihat pada Gambar 4.



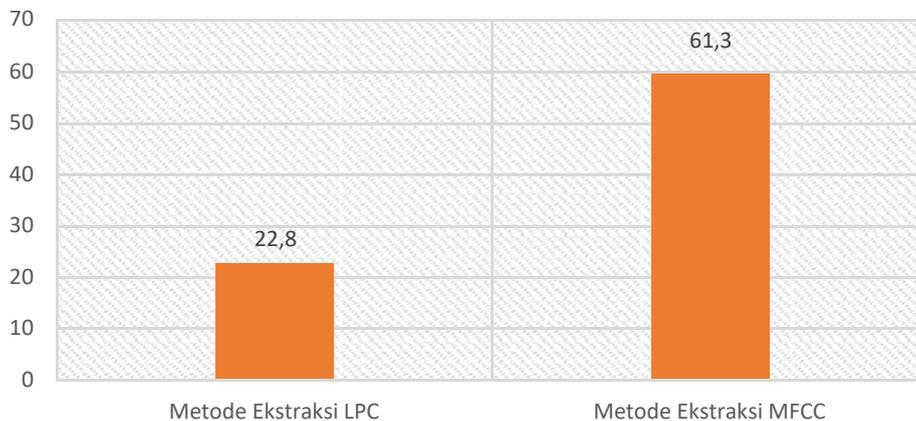
Gambar 4. Hasil Pengenalan Emosi Menggunakan Ekstraksi MFCC dan Klasifikasi MLP

D. Perbandingan Hasil Pengenalan Emosi

Berdasarkan penelitian yang telah dilakukan pada metode ekstraksi fitur LPC dan MFCC dengan klasifikasi Jaringan Syaraf Tiruan (MLP), perbandingan akurasinya

dapat dilihat pada Gambar 5. Hasil persentase didapatkan dari penjumlahan persentase lima emosi yang kemudian dicari nilai rata-ratanya.

Perbandingan Pengenalan Emosi Metode LPC dan MFCC



Gambar 5. Perbandingan Pengenalan Emosi Metode LPC dan MFCC

Berdasarkan penelitian yang dilakukan pada dua metode ekstraksi fitur, hasil akurasi pengenalan emosi untuk ekstraksi fitur LPC dan klasifikasi Jaringan Syaraf Tiruan (MLP) adalah 22,8% dan hasil pengenalan emosi untuk ekstraksi fitur MFCC adalah 61,33%. Masing-masing metode menggunakan data emosi marah, bosan, bahagia, netral, dan sedih. Data tersebut dibagi menjadi dua, yaitu data *testing* dan data *training* dengan perbandingan 80:20. Kedua metode tersebut sama-sama menggunakan tiga lapisan yaitu lapisan input, lapisan tersembunyi, dan lapisan output.

KESIMPULAN

Pengenalan ucapan emosi suara dengan membandingkan metode ekstraksi fitur LPC dan metode ekstraksi fitur MFCC yang masing-masing metode diklasifikasikan dengan metode Jaringan Syaraf Tiruan (MLP) menghasilkan nilai akurasi yang masih rendah. Hasil akurasi untuk metode ekstraksi fitur LPC sebesar 22,8% pada semua emosi. Sedangkan Hasil akurasi untuk metode ekstraksi fitur MFCC sebesar 61,33% pada semua emosi. Pada hasil klasifikasi dengan dua metode ekstraksi fitur yang berbeda, akurasi pengenalan emosi tertinggi adalah emosi netral yaitu 28% untuk ekstraksi fitur LPC dan 71% untuk ekstraksi fitur MFCC. Hasil akurasi ini bisa ditingkatkan dengan menambahkan data yang lebih banyak lagi, terutama untuk data *testing*. Perlunya pengujian pada nilai parameter jaringan MLP, yaitu dengan mengubah nilai-nilai parameter, seperti learning rate, epsilon, epoch, dan Cross Validation, karena dapat mempengaruhi tingkat akurasi pengenalan. Selain itu penentuan ekstraksi fitur dan klasifikasi metode yang lain juga dapat digunakan untuk mencari nilai akurasi pengenalan emosi yang lebih baik lagi.

UCAPAN TERIMAKASIH

Riset ini didukung dan didanai oleh Kemenristek Dikti dengan Nomer Kontrak: PTM-026/SKPP.TT/LPPM UAD/111/2019.

DAFTAR PUSTAKA

- [1] S. Vaishnav and S. Mitra, "Speech Emotion Recognition: A Review," *Int. Res. J. Eng. Technol. IRJET*, vol. 3, no. 04, pp. 313–316, 2016.
- [2] S. Lalitha, A. Madhavan, B. Bhushan, and S. Saketh, "Speech Emotion Recognition," *2014 Int. Conf. Adv. Electron. Comput. Commun. ICAECC 2014*, vol. 7, 2015, doi: 10.1109/ICAIECC.2014.7002390.
- [3] A. B. Gumelar, "Human Voice Emotion Identification Using Prosodic and Spectral Feature Extraction Based on Deep Neural Networks," *2019 IEEE 7th Int. Conf. Serious Games Appl. Health SeGAH*, pp. 1–8, 2019.
- [4] M. D. Pell and S. A. Kotz, "On the time course of vocal emotion recognition," *PLoS One*, vol. 6, no. 11, p. e27256, 2011.
- [5] I. Idrisa, M. S. H. Salamb, and M. S. Sunarc, "Speech Emotion Classification Using SVM and MLP on Prosodic and Voice Quality Features," *J. Teknol.*, vol. 78, 2015.
- [6] G. Liu, W. He, and B. Jin, "Feature Fusion of Speech Emotion Recognition Based Deep Learning," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, 2018, pp. 193–197, doi: 10.1109/ICNIDC.2018.8525706.
- [7] Y. Sun, G. Wen, and J. Wang, "Weighted spectral features based on local Hu moments for speech emotion recognition," *Biomed. Signal Process. Control*, vol. 18, pp. 80–90, 2015, doi: 10.1016/j.bspc.2014.10.008.
- [8] S. S. Swaminathan and J. Thangaiyan, "Emotion Speech Recognition using MFCC and Residual Phase in Artificial Neural Network," *Int. J. Eng. Res. Sci. Technol.*, vol. 4 No.3, no. August, pp. 106–113, 2015.
- [9] Irmawan, H. Hikmarika, D. W. Sari, and M. C. Tammimi, "Pengenalan Kata dengan Metode Linear Predictive Coding dan Jaringan Syaraf Tiruan Pada Mobile Robot," in *Conference on Information Technology and Electrical Engineering*, 2014, no. October 2014, pp. 139–144.
- [10] K. V. Krishna Kishore and P. Krishna Satish, "Emotion Recognition in Speech using MFCC and Wavelet Features," *Proc. 2013 3rd IEEE Int. Adv. Comput. Conf. IACC 2013*, pp. 842–847, 2013, doi: 10.1109/IAdCC.2013.6514336.