ISSN 2527-3671 (*Online*)
ISSN 2087-8893 (*Print*)

Teknologi: Jurnal Ilmiah Sistem Informasi 13 (2) July 2023 1-11

Tersedia *online* di www.journal.unipdu.ac.id

Unipdu

Halaman jurnal di www.journal.unipdu.ac.id/index.php/teknologi

Klasifikasi Jenis Emosi Melalui Ucapan Menggunakan Metode Convolutional Neural Network

Fransiskus Jonatan Tanudjaja [a], Eva Yulia Puspaningrum [b], Yisti Vita Via [c]

[a, b, c] Program Studi Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia

email: [a,*] *19081010062@student.upnjatim.ac.id*

*Korespondensi

Dikirim 05 Juni 2023; **Direvisi** 08 Juni 2023; **Diterima** 20 Juni 2023; **Diterbitkan** 10 Juli 2023

Abstrak

Ucapan adalah metode yang paling sering digunakan manusia untuk berkomunikasi satu sama lain. Ucapan berisi informasi yang bervariasi dimana selain dapat mengetahui pesan seseorang, juga dapat mengetahui kondisi emosi orang tersebut. Ekspresi emosi dalam sebuah percakapan berperan penting dalam memberikan penekanan pada informasi yang disampaikan menjadi lebih kuat. Pengenalan emosi melalui ucapan ini dapat diaplikasikan ke dalam berbagai bidang seperti ilmu kognitif, *call centre* dan bidang lainnya. Oleh karena itu, penelitian ini bermaksud untuk mengklasifikasikan emosi seseorang melalui ucapan mereka dengan menggunakan algoritma *Convolutional Neural Network* (CNN). Tujuan penelitian ini adalah untuk mencari model CNN dengan performansi terbaik dalam mengklasifikasikan emosi menjadi 8 kelas yaitu, netral, sedih, tenang, takut, senang, terkejut, jijik dan marah. Model CNN dibedakan berdasarkan data masukkan yang menggunakan metode ekstraksi fitur yang berbeda-beda. Dari hasil pengujian diperoleh model paling baik dengan rasio pembagian data sebesar 80% untuk data latih, 10% untuk data validasi serta 10% untuk data uji yang memakai metode ekstraksi *Mel-Frequency Cepstral Coefficients* (MFCC) dengan nilai rata-rata akurasi sebesar 70% diikuti nilai rata-rata *recall* dan presisi masing-masing 68% dan 67%. Untuk emosi yang paling sering ditebak dengan benar adalah emosi marah, terkejut, sedih dan tenang dengan rata-rata prediksi benar sebesar 77%.

Kata Kunci: Convolutional Neural Network, klasifikasi emosi, Mel-Frequency Cepstral Coefficients, ucapan

Type Of Emotions Classification Based On Speech Using Convolutional Neural Network Method

Abstract

Speech is the method most often used by humans to communicate with other. Speech contains varied informations. In addition to being able to know someone's message, it can also find out the person's emotional condition. The expression of emotion in a conversation plays an important factor in emphasizing the information conveyed. Recognition of emotions through speech can be applied to various fields such as cognitive science, call centers, and other fields. Therefore, Convolutional Neural Network (CNN) algorithm will be used to classify a person's emotions through their speech. The main objective of this study was to find CNN model with the best performance in classifying emotions into 8 classes: calm, neutral, sad, happy, scared, angry, surprised, and disgusted. CNN model is distinguished based on input data using different feature extraction methods. From the test results, the best model is obtained with data sharing ratio of 80% for training data, 10% for validation data, and 10% for test data using Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method with an average accuracy value of 70%, followed by the average values of precision and recall of 68% and 67%, respectively. The emotions that were most often correctly guessed were anger, surprise, sadness, and calm, with an average correct prediction of 77%.

Keywords: Convolutional Neural Network, emotion classification, Mel-Frequency Cepstral Coefficients, speech

Untuk mengutip artikel ini dengan *APA Style*:

Tanudjaja. J. F, Puspaningrum. E. Y, Via. Y.F (2023). Klasifikasi Jenis Emosi Melalui Ucapan Menggunakan Metode Convolutional Neural Network. TEKNOLOGI: Jurnal Ilmiah Sistem Informasi, 13(2), 1-11: https://doi.org/10.26594/teknologi.v13i2.3740

1. Pendahuluan

Ucapan adalah metode yang disukai dan paling sering dipakai manusia untuk berkomunikasi antara satu dengan yang lain. Karena dengan ucapan manusia dapat membagikan informasi dan perasaannya dengan lebih cepat (Jain, et al., 2018). Ucapan itu sendiri berisi informasi yang bervariasi dimana selain dapat mengetahui pesan dan maksud seseorang, kita juga dapat mengetahui kondisi emosi orang tersebut (Helmiyah, et al., 2020). Emosi sendiri merupakan bentuk ekspreksi manusia terhadap suatu hal ataupun seseorang. Ekspresi emosi dalam sebuah percakapan berperan penting dalam memberikan penekanan

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

pada informasi yang disampaikan menjadi lebih kuat daripada hanya dengan kata-kata saja (Lasiman & Lestari, 2018). Hal ini menyebabkan penelitian untuk mendeteksi emosi manusia melalui sebuah ucapan sedang gencar dilakukan. Pengenalan emosi melalui sebuah ucapan merupakan suatu tindakan untuk mengenali emosi seseorang serta keadaan afektif pada sebuah percakapan. Hal ini didasarkan pada fakta dimana setiap suara yang dikeluarkan mencerminkan emosi melalui nada dan intonasi.

Pengenalan emosi melalui ucapan ini sangat berguna dan dapat diaplikasikan ke dalam berbagai bidang seperti ilmu kognitif, psikologi, *neuroscience, call centre*, industri game, kesehatan dan bidang lainnya (Jain, et al., 2018). Salah satu contoh penerapannya adalah pada kokpit pesawat, dimana digunakan untuk mengetahui kondisi mental dari pilot sehingga dapat menghindari terjadinya kecelakaan. Selain itu juga pada percakapan call centre, untuk dapat mengetahui kondisi mood seorang pelanggan sehingga dapat digunakan untuk meningkatkan kualitas pelayanan *call centre* tersebut. Pesatnya perkembangan teknologi informasi saat ini yang mana tidak lepas mengenai bidang pemrosesan sinyal digital. Memungkinkan perancangan suatu sistem cerdas yang dapat mengidentifikasi jenis emosi manusia melalui ucapan secara otomatis. Berdasarkan hal ini, penulis membuat sistem yang dapat mengidentifikasi jenis emosi manusia berdasarkan ucapan dengan menggunakan metode *Convolutional Neural Network*.

Berdasarkan studi literatur yang dilakukan penulis terbukti metode *convolutional neural network* ini cocok dan memberikan hasil akurasi yang baik dalam memprediksi kelas emosi seseorang dikarenakan adanya kompleksitas pada fiturnya (Jain, et al., 2018). Penelitian lainnya yang serupa dilakukan oleh (Zhao, et al., 2019; Qayyum, et al., 2019) dimana didapatkan hasil penggunaan dengan metode CNN untuk identifikasi emosi memberikan hasil akurasi paling tinggi jika dibandingkan dengan penggunaan metode lainnya seperti *Support Vector Machine (SVM), Multivariate Linear Regression Classification (MLR)* dan *Recurrent Neural Network (RNN)* yaitu sebesar 83.61%.

Penelitan lainnya yang dilakukan oleh Yulistia Khoirotul Aini beserta rekannya yang berjudul "Pemodelan CNN Untuk Deteksi Emosi Berbasis *Speech* Bahasa Indonesia" (Aini, et al., 2021). Dataset yang dipakai pada penelitian ini berjumlah 507 *file audio* yang diambil dari *TV series* berjudul "*Imperfect*". Pada penelitian ini, *dataset file audio* yang telah dikumpulkan akan diklasifikasikan ke dalam 4 kelas emosi, yaitu emosi marah, netral, senang dan sedih. Selanjutnya dipakai metode *Mel-Frequency Cepstrum Coefficient (MFCC),* frekuensi fundamental serta *Root Mean Square Energy (RMSE)* untuk tahapan ekstraksi fiturnya. Setelah itu dilanjutkan dengan melakukan klasifikasi dengan menggunakan algoritma *Convolutional Neural Network*. Dari skenario percobaan yang telah dilakukan didapatkan hasil terbaik dengan melakukan kombinasi fitur MFCC dan *pitch* dengan hasil rata-rata akurasi mencapai 85%.

Penelitian serupa dilakukan oleh (Zhao, et al., 2019) dengan judul "*Speech emotion recognition using deep 1D & 2D CNN LSTM networks*". Pada penelitian ini digunakan 2 dataset yaitu Berlin EmoDB dan *Interactive Emotional Dyadic Motion Capture (IEMOCAP)* yang masing-masing berjumlah 535 dan 1150 *file audio. File audio* pada dataset tersebut akan diklasifikasikan menjadi 6 kelas emosi yang terdiri dari emosi marah, bersemangat, kecewa, senang, netral dan sedih. Penelitian ini memakai algoritma *Long Short Term Memory (LSTM)*. Berdasarkan penelitian yang telah dilakukan oleh penulis didapatkan hasil terbaik dari model yang menggunakan LSTM 2 dimensi dengan hasil akurasinya sebesar 95.89% untuk dataset Berlin EmoDB dan 52.14% untuk dataset IEMOCAP.

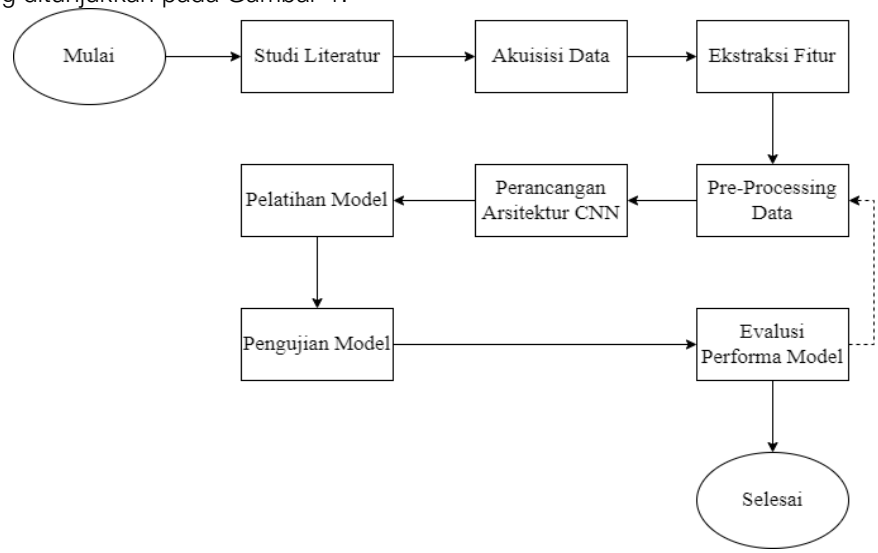
Penelitian terkait juga pernah dilakukan oleh Udit Jain beserta rekannya pada tahun 2018 yang berjudul "*Cubic SVM Classifier Based Feature Extraction and Emotion Detection from Speech Signals*" (Jain, et al., 2018). Dataset yang digunakan dalam penelitian ini dikumpulkan dari unduhan *file audio* 20 artis Bollywood pada youtube yang berjumlah 400 file audio. Pada penelitian ini, emosi pada setiap file audio akan diklasifikasikan menjadi 4 kelas yaitu emosi senang, sedih, marah dan netral. Tahapan ekstraksi fitur digunakan beberapa metode seperti *energy, pitch, formant, Mel-Frequency Cepstrum Coefficient (MFCC), dan Linear Prediction Cepstral Coefficient (LPCC).* Pada penelitian ini penulis menggunakan algoritma *cubic spine Support Vector Machine (SVM)* untuk mengklasifikasikan jenis emosi yang ada. Dari beberapa skenario pengujian yang dilakukan hasil terbaiknya didapatkan dengan nilai akurasi sebesar 98.75%, presisi sebesar 97.73%, *recall* sebesar 97.5% dan *f1-score* sebesar 97.5%.

Berdasarkan penelitian sebelum-sebelumnya didapati bahwa kebanyakan penelitian sebelumnya hanya dapat mengklasifikasikan emosi ke dalam beberapa kelas umum saja. Oleh sebab itum, pada penelitian ini penulis bertujuan untuk membuat model CNN yang dapat mengklasifikasikan lebih banyak jenis emosi dibandingkan penelitian sebelumnya. Terdapat 8 jenis emosi yang dapat diklasifikasikan oleh model yang akan dibuat antara lain, emosi tenang, netral, sedih, senang, takut, marah, terkejut, dan jijik.

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

2. Metode Penelitian

Pada penelitian ini terdapat beberapa langkah yang harus dilakukan untuk dapat mengklasifikasikan jenis emosi melalui ucapan menggunakan metode *convolutional neural network*. Berikut ini tahapan penelitian yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Studi Literatur

Tahapan pertama yang dilakukan adalah mengumpulkan literatur terkait topik penelitian yang sedang diangkat. Tahapan ini berfungsi untuk memperoleh penjelasan mengenai teori-teori serta konsep dari metode serta permasalahan yang diteliti, seperti pada topik penelitian ini penulis membaca literatur mengenai *speech emotion recognition*, algoritma *convolutional neural network*, dan lainnya. Selain itu juga bertujuan untuk mengetahui gap dari penelitian yang telah lalu.

2.2. Akuisisi Data

Data yang dipakai pada topik penelitian ini adalah suara manusia dalam bahasa inggris. Data ini sendiri sudah memiliki label sebelumnya. *The Ryerson Audio-Visual Dataset of Emotional Speech and song (RAVDESS) Dataset* ini dikumpulkan oleh penulis dari website kaggle dengan tautan sebagai berikut (https://www.kaggle.com/uwrfkaggler/RAVDESS-emotional-speech-audio). Aksen suara yang ada pada *dataset* merupakan aksen Amerika Utara. Dataset ini sendiri terdiri dari dua macam file yaitu file yang berasal dari percakapan seseorang serta file yang berisikan lagu. Pada penelitian ini penulis akan menggunakan file dari rekaman percakapan yang berjumlah 1440 *file audio*. Setiap *audio* pada dataset ini memiliki durasi antara 3,5 detik sampai dengan 4 detik. Rekaman percakapan pada *dataset* ini terdiri dari suara laki-laki dan perempuan dengan jumlah masing-masingnya 12 orang. Format setiap *file audio* pada *dataset* yang dipakai pada penelitian ini adalah 16 bit, 48 kHz dengan tipe .wav.

2.3. Ekstraksi Fitur

Tahapan selanjutnya adalah mengekstraksi fitur-fitur pada tiap-tiap *file audio* yang ada pada *dataset*. Pada proses ekstraksi fitur ini, peneliti memakai beberapa metode yang akan dijelaskan pada Tabel 1.

Tabel 1. Metode ekstraksi yang dipakai

Metode Ekstraksi Fitur	Keterangan
Mel-Frequency Cepstral Coefficients	Mendeteksi perbedaan frekuensi antar suara dengan tepat.
Energy	Menghitung jumlah energi pada sinyal audio.
Pitch	Mengukur jumlah rata-rata osilasi per detik.
Spectral Centroid	Mengukur spektral komponen pada sinyal suara yang membantu dalam membedakan emosi dalam sebuah suara.
Spectral Flatness	Mengukur kuantitas noise pada sebuah sinyal audio.
Spectral Roll-off	Mengukur arah ketidaksimetrisan pada sebuah spektrum dalam sinyal audio.

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

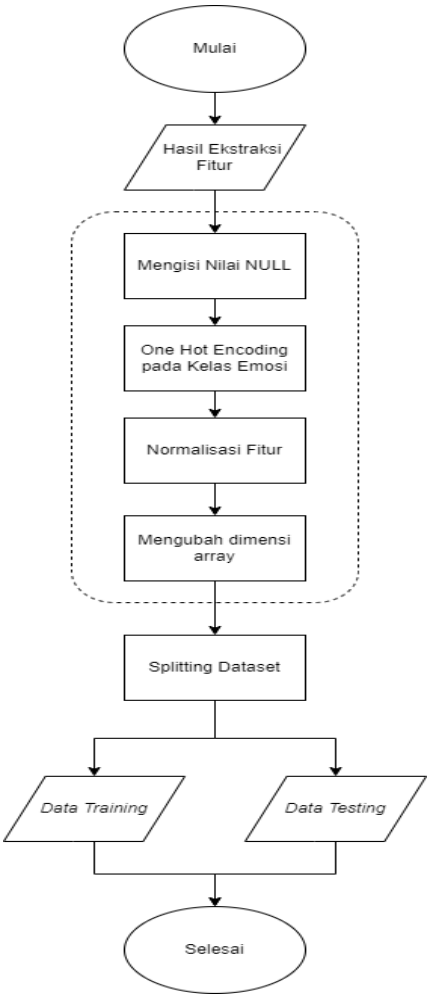
	0	1	2	3	4	5	6	7	8	9	...	161	162	163	164	165	166	167	168	169	labels
0	-556.887085	65.237015	-0.884413	15.528003	9.315155	-3.457776	-4.759561	-9.590000	-14.681817	-2.406363	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	takut
1	-666.618530	75.513725	0.754220	15.570239	6.638893	3.128690	-1.838934	-5.840415	-10.310457	-5.500113	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tenang
2	-660.230347	63.325817	-2.630458	17.983356	9.407701	-0.981498	-2.324698	-6.848032	-14.133670	-2.200332	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	netral
3	-329.682037	37.055607	-18.648388	6.549638	-9.696449	-6.251557	-13.365948	-8.641981	-21.028406	-8.458943	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	marah
4	-707.943298	70.332085	-3.981637	16.394806	7.665534	-2.153185	-2.060916	-8.838734	-12.730490	-4.445475	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sedih
5	-657.722351	65.035187	3.148672	15.666512	7.582659	2.628088	-1.850068	-7.477580	-11.976264	-4.458464	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	netral
6	-681.650208	69.696854	0.469155	18.170815	6.870313	-0.448264	1.356428	-8.938222	-14.520138	-4.707395	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sedih
7	-475.010406	64.748047	-11.656610	8.972133	-1.362460	-4.326521	-12.854100	-12.519542	-22.508835	-9.462560	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	senang
8	-460.330261	63.159615	-13.468574	9.937035	-0.075396	-3.240332	-13.818613	-12.049685	-23.861624	-6.467387	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	senang
9	-607.864075	69.669411	-0.695821	21.833481	2.061868	-0.178393	-3.930124	-8.634003	-17.164713	-5.271435	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	senang
10	-327.228424	38.104771	-20.152468	-1.265623	-8.632886	-4.885595	-13.674493	-19.788263	-20.102552	-11.717583	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	takut
11	-676.023438	80.262627	2.503204	13.308867	11.569080	1.193359	-5.365985	-5.692616	-9.998198	-5.729336	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tenang
12	-694.579712	72.531723	3.104566	17.112118	9.077340	1.836257	-4.146086	-4.797150	-11.764124	-5.925480	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tenang
13	-587.739807	62.540543	-6.062552	12.954743	4.547711	-4.493005	-10.604271	-8.765388	-18.683168	-5.465188	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sedih
14	-624.142578	73.200943	-2.984951	15.253687	2.033356	-0.278258	-6.654721	-8.612391	-16.168873	-3.934198	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	senang
15	-601.020020	73.714249	-7.142501	18.018578	7.663852	-1.983695	-7.730342	-5.682049	-16.004137	-1.610288	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	senang
16	-667.117126	79.265198	0.214563	14.969451	8.738843	1.614774	-1.532765	-7.318048	-9.807106	-5.784626	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tenang
17	-577.984131	67.760506	-7.037594	16.876732	4.939756	-3.477778	-8.413762	-11.454423	-22.813313	-5.983521	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sedih
18	-520.926086	49.648300	-6.148461	8.639017	0.098176	-3.918394	-7.859183	-8.987469	-13.624468	-7.812575	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	terkejut
19	-687.393799	86.303032	4.910166	16.588545	11.206296	2.007060	-4.878830	-3.813346	-7.948537	-5.705545	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	tenang

Gambar 2. Dataframe hasil ekstraksi fitur beserta kelas emosi

Gambar 2 menunjukkan fitur suara yang telah selesai diekstraksi dan dibentuk ke dalam sebuah dataframe bernama 'hasil_ekstraksi, untuk memudahkan pengolahan pada proses selanjutnya.

2.4. *Pre-processing Data*

Tahapan selanjutnya adalah melakukan pra-proses pada fitur yang telah selesai diekstrak. Tahapan pre-processing ini berguna untuk menghindari permasalahan yang mungkin terjadi pada saat tahapan klasifikasi data nantinya dikarenakan data belum sesuai. Langkah-langkah pre-processing data dijelaskan menggunakan diagram alir pada Gambar 3.

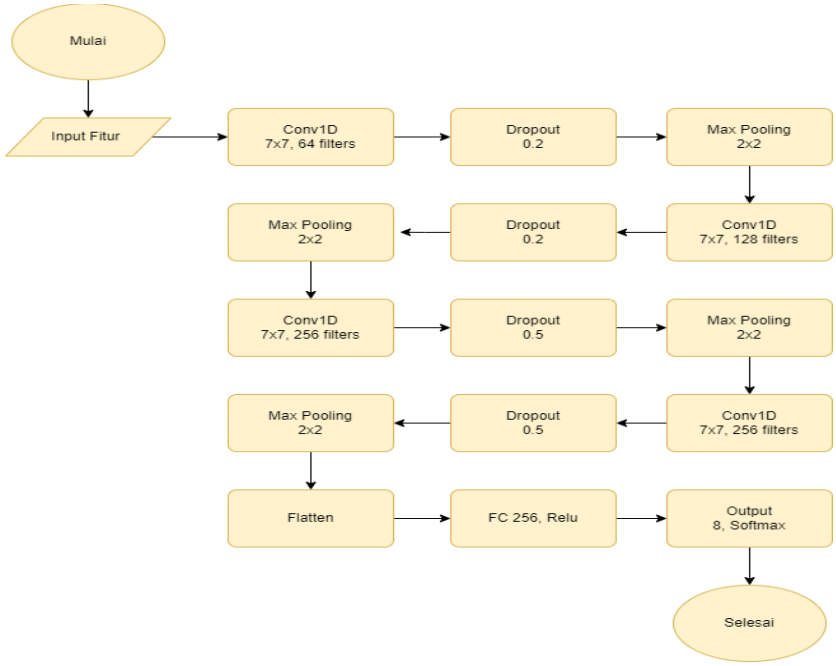


Gambar 3. Tahapan *pre-processing data*

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

2.5. Perancangan Arsitektur CNN

Tahapan selanjutnya adalah melakukan perancangan arsitektur untuk model *convolutional neural network*. Pada Gambar 4 dijelaskan rancangan CNN yang akan dipakai untuk penelitian ini. Arsitektur CNN pada penelitian ini terdiri dari 4 lapisan konvolusi diikuti dengan lapisan *fully connected*. Pada setiap lapisan konvolusi terdiri dari, layer konvolusi dengan fungsi aktivasi relu yang berguna untuk memproses *output* dari proses konvolusi matrix, dropout untuk mengabaikan sebagian neuron pada matrix dan max pooling untuk mengurangi dimensi matriks hasil konvolusi. Selanjutnya pada lapisan *fully connected* terdiri dari *flatten* yang berguna untuk mengubah dimensi *output* hasil konvolusi menjadi satu dimensi agar dapat dilakukan klasifikasi. Setelah itu akan masuk ke dalam hidden layer dan *output* (dense). Pada layer *output* digunakan fungsi aktivasi softmax agar dapat mengklasifikasikan kelas emosi menjadi 8 kelas yang ada.



Gambar 4. Rancangan arsitektur CNN

2.6. Pelatihan dan Pengujian Model

Tahapan selanjutnya adalah pelatihan model *convolutional neural network* dengan *data training* yang telah melalui proses *pre-processing* sebelumnya. Sebelum dilakukannya pelatihan model CNN, diperlukan inisialisasi beberapa parameter yang diperlukakan seperti fungsi loss, fungsi optimizer, jumlah epoch dan callback. Berikut ini inisialisasi parameter untuk model CNN ditunjukkan pada Tabel 2.

Tabel 2. Inisialisasi Parameter CNN

Parameter	Masukkan
Output Class	8
Output Layer Activation	Softmax
Hidden Layer Activation	Relu
Epoch	50
Batch Size	64
Optimizer Function	Adam
Loss Function	Categorical cross entropy

Kemudian setelah parameter-parameter tersebut diinisialisasi selanjutnya data training akan dilatih sesuai dengan arsitektur CNN yang ada. Pada saat proses pelatihan ini juga akan dilakukan validasi dengan data validasi untuk melihat apakah model yang dilatih data training benar-benar belajar atau tidak. Hasil dari pelatihan ini merupakan sebuah model *convolutional neural network* yang terlatih. Setelah tahapan pelatihan selesai dilakukan maka akan dilanjutkan ke tahap pengujian model *convolutional neural network* yang sudah terlatih, pengujian dilakukan menggunakan *data testing* yang didapatkan setelah melakukan pembagian data dengan rasio tertentu. Proses pengujian ini dilakukan dengan membandingkan hasil prediksi dengan label aslinya yang terdapat pada *data testing* untuk melihat akurasi prediksi model tersebut. Setelah itu akan dilanjutkan ke tahapan evaluasi model menggunakan *confussion matrix*.

Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11
ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

2.7. Evaluasi Performa Model

Tahapan selanjutnya adalah melakukan evaluasi pada model yang ada. Evaluasi model ini bertujuan untuk mengetahui seberapa optimal performa dari model yang ada. Evaluasi model dilakukan dengan menggunakan *data testing* menggunakan model CNN yang terlatih dengan metode *confussion matrix* untuk mendapatkan nilai *f1-score*, *recall,* presisi, serta akurasi.

3. Hasil dan Pembahasan

3.1. Rasio Pembagian Data

Pembagian data dilakukan menggunakan beberapa skenario pengujian. Hal ini bertujuan untuk mendapatkan rasio pembagian data terbaik. Rasio pembagian data dibagi menjadi 4 skenario yaitu, 90:5:5, 80:10:10, 70:15:15 dan 60:20:20 dengan berturut-turut *data training, data validasi* dan *data testing*. Hal ini diterapkan pada data yang berisi fitur MFCC saja untuk mengurangi waktu komputasi. Hasil dari percobaan yang dilakukan ditunjukkan pada Tabel 3.

Tabel 3. Perbandingan akurasi untuk rasio pembagian data yang berbeda-beda

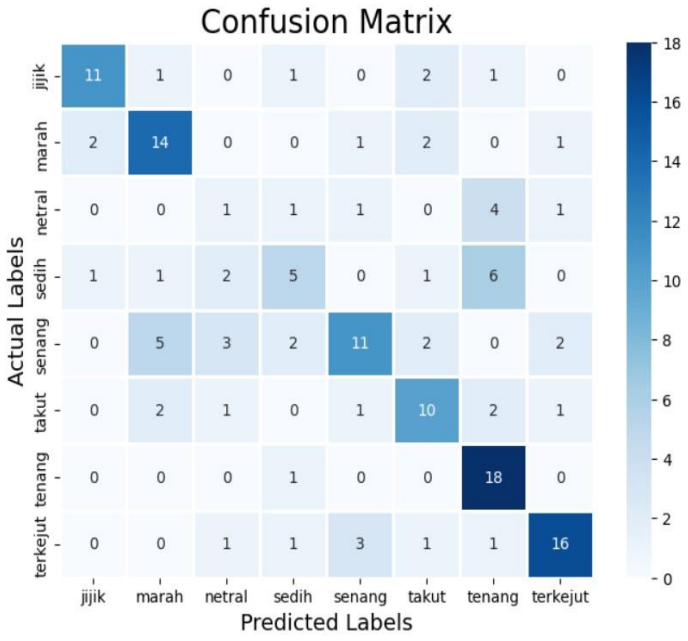
Data Latih (%)	Data Validasi (%)	Data Uji (%)	Akurasi Latih (%)	Akurasi Validasi (%)
90	5	5	94	66
80	10	10	92	74
70	15	15	92	70
60	20	20	91	63

Pada Tabel 3 terlihat bahwa pembagian data dengan rasio 80% untuk data latih, 10% untuk data validasi dan 10% untuk data uji memperoleh nilai akurasi pelatihan dan validasi yang lebih baik dibandingkan dengan rasio pembagian data lainnya.

3.2. Skenario Pengujian

a) Pengujian menggunakan metode ekstraksi *MFCC, pitch, energy* dan *spectral components*

Dari hasil pengujian yang ditunjukkan pada Gambar 5 diambil kesimpulan bahwa cukup banyak kelas yang masih salah ditebak, terutama untuk emosi netral dengan hanya 1 data yang diprediksi dengan benar karena masih banyak yang tertebak ke emosi tenang serta emosi sedih dengan 5 data diprediksi dengan benar karena banyak yang tertebak ke emosi tenang. Selain itu untuk kelas yang paling banyak diprediksi dengan benar adalah emosi terkejut dengan banyak data yang ditebak dengan benar sebanyak 16 data dari keseluruhan 23 data serta emosi tenang dengan total data yang ditebak dengan benar sebanyak 18 dari keseluruhan 19 data.



Gambar 5. *Confusion matrix* untuk pengujian menggunakan semua metode ekstraksi fitur

Berdasarkan Tabel 4 untuk skema pengujian dengan menggunakan semua metode ekstraksi fitur dihasilkan akurasi rata-rata untuk keseluruhan kelas emosi sebesar 60%. Serta nilai rata-rata masing-masing *f1-score, recall* dan presisi adalah 55%, 56%, dan 56%.

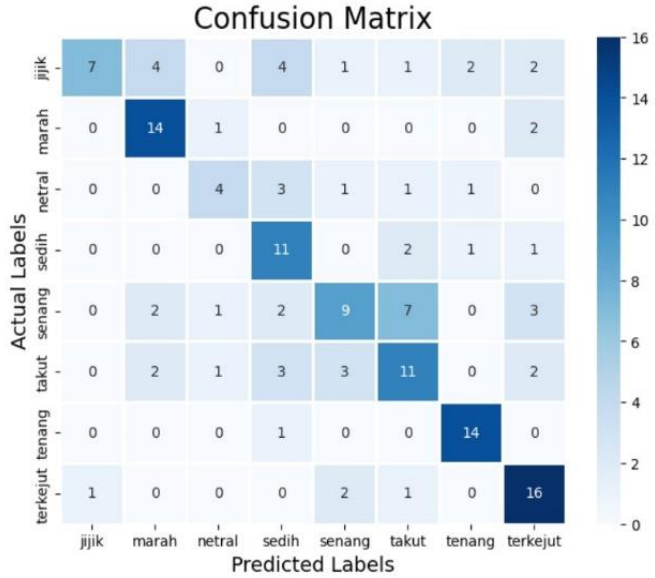
ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

Tabel 4. Perbandingan nilai akurasi, *f1-score*, presisi dan *recall* dengan menggunakan semua metode ekstraksi

Kelas Emosi	F1-score (%)	Presisi (%)	Recall (%)	Rata-rata Akurasi (%)
Jijik	73	79	69	60
Marah	65	61	70	
Netral	12	12	12	
Sedih	37	45	31	
Senang	52	65	44	
Takut	57	56	59	
Tenang	71	56	95	
Terkejut	73	76	70	

b) Pengujian menggunakan metode ekstraksi *MFCC, pitch* dan *energy*

Dari Gambar 6 ditarik kesimpulan terhadap hasil uji model bahwa masih cukup banyak kelas yang tertebak salah, terutama untuk kelas jijik dengan hanya memiliki 7 data diprediksi dengan benar, netral dengan total 4 data diprediksi dengan benar dan senang dengan total 9 data diprediksi dengan benar. Sedangkan untuk kelas yang paling banyak berhasil diprediksi dengan benar adalah kelas marah dengan banyak data yang ditebak dengan benar sebanyak 14 data dari keseluruhan 17 data, emosi terkejut dengan banyak data yang ditebak dengan benar sebanyak 16 dari total 20 data yang ada dan emosi tenang dengan banyak data yang ditebak dengan benar sebanyak 14 dari keseluruhan 15 data yang ada. Alasannya masih banyaknya kelas yang salah diprediksi adalah karena karakter suara untuk beberapa kelas sangatlah mirip dengan kelas lainnya sehinnga model tidak dapat membedakan dengan benar.



Gambar 6. *Confusion matrix* untuk pengujian menggunakan metode *MFCC, pitch* dan *energy*

Berdasarkan Tabel 5 untuk skema pengujian dengan menggunakan metode ekstraksi metode *MFCC, pitch* dan *energy* dihasilkan akurasi rata-rata untuk keseluruhan kelas emosi sebesar 60%. Serta nilai rata-rata masing-masing *f1-score, recall* dan presisi adalah 59%, 61%, dan 62%.

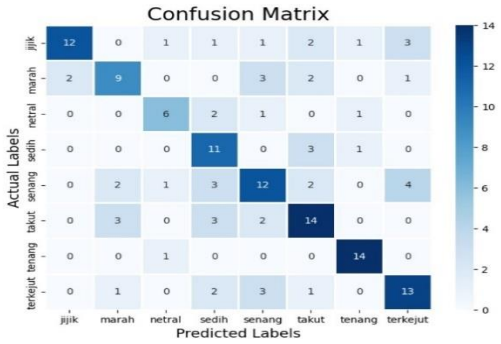
Tabel 5. Perbandingan nilai akurasi, *f1-score*, presisi dan *recall* dengan menggunakan *MFCC, pitch* dan *energy*

Kelas Emosi	F1-score (%)	Presisi (%)	Recall (%)	Rata-rata Akurasi (%)
Jijik	48	88	33	60
Marah	72	64	82	
Netral	47	57	40	
Sedih	56	46	73	
Senang	45	56	38	
Takut	49	48	50	
Tenang	85	78	93	
Terkejut	70	62	80	

Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11
ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

c) Pengujian menggunakan metode ekstraksi *MFFC* dan *energy*

Dari Gambar 7 diambil kesimpulan terhadap hasil uji model dimana model sudah mendapatkan hasil yang cukup baik karena hampir semua kelas sudah dapat diprediksi dengan benar. Namun masih terdapat kelas yang masih sering salah prediksi yaitu kelas senang dengan hanya memiliki 12 data yang ditebak dengan benar dari keseluruhan 24 data yang ada, serta kelas netral dengan 6 data yang ditebak dengan benar dari keseluruhan 10 data. Sedangkan kelas emosi yang paling banyak berhasil diprediksi dengan benar adalah kelas tenang dengan banyak data yang ditebak dengan benar sebanyak 14 data dari keseluruhan 15 data yang ada.



Gambar 7. Confusion matrix untuk pengujian menggunakan metode MFCC dan energy

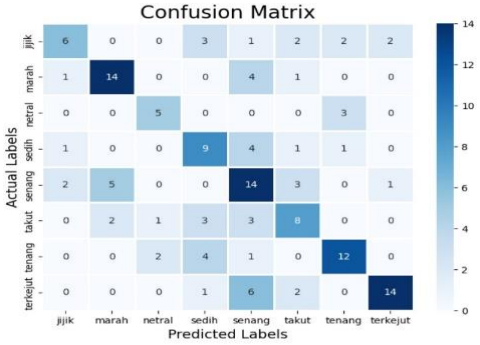
Berdasarkan Tabel 6 untuk skema pengujian dengan menggunakan metode ekstraksi *MFCC* dan *energy* dihasilkan akurasi rata-rata untuk keseluruhan kelas emosi sebesar 63%. Serta nilai rata-rata masing-masing *f1-score, recall* dan presisi adalah 64%, 64%, dan 65%.

Tabel 6. Perbandingan nilai akurasi, *f1-score,* presisi dan *recall* menggunakan *MFCC* dan *energy*

Kelas Emosi	F1-score (%)	Presisi (%)	Recall (%)	Rata-rata Akurasi (%)
Jijik	69	86	57	63
Marah	56	60	53	
Netral	63	67	60	
Sedih	59	50	73	
Senang	52	55	50	
Takut	61	58	64	
Tenang	87	82	93	
Terkejut	63	62	65	

d) Pengujian menggunakan metode ekstraksi *MFFC* dan *pitch*

Pada Gambar 8 dapat diambil kesimpulan terhadap hasil uji model dimana banyak kelas yang masih salah ditebak jika dibandingkan dengan skenario pengujian lainnya skenario dengan menggunakan gabungan metode *MFCC* dan *pitch* ini merupakan skenario yang mendapatkan hasil yang paling buruk. Terutama untuk kelas emosi yang masih sering salah prediksi adalah kelas jijik dengan hanya memiliki 6 data yang ditebak dengan benar dari keseluruhan 16 data yang ada, serta kelas takut dengan 8 data yang ditebak dengan benar dari keseluruahn 17 data. Sedangkan kelas emosi yang cukup baik dalam hasil prediksinya adalah kelas terkejut dengan banyak data yang ditebak dengan benar sebesar 14 data dari 23 data yang ada.



Gambar 8. *Confusion matrix* untuk pengujian menggunakan metode *MFCC* dan *pitch*

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)

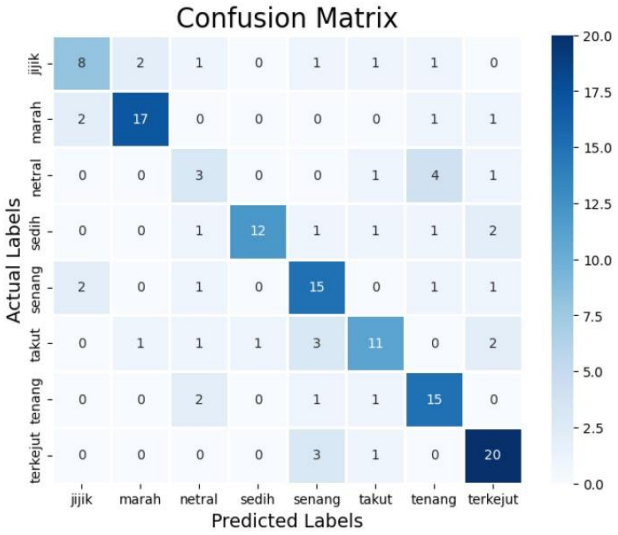
Berdasarkan Tabel 7 untuk skema pengujian dengan menggunakan metode ekstraksi *MFCC* dan *pitch* dihasilkan akurasi rata-rata untuk keseluruhan kelas emosi sebesar 57%. Serta nilai rata-rata masing-masing *f1-score, recall* dan presisi adalah 57%, 57%, dan 59%.

Tabel 7. Perbandingan nilai akurasi, *f1-score,* presisi dan *recall* menggunakan *MFCC* dan *Pitch*

Kelas Emosi	F1-score (%)	Presisi (%)	Recall (%)	Rata-rata Akurasi (%)
Jijik	46	60	38	57
Marah	68	67	70	
Netral	62	62	62	
Sedih	50	45	56	
Senang	48	42	56	
Takut	47	47	47	
Tenang	65	67	63	
Terkejut	70	82	61	

e) Pengujian menggunakan metode ekstraksi *MFFC*

Pada Gambar 9 dapat diambil kesimpulan terhadap hasil uji model dimana model dengan skenario ekstraksi fitur menggunakan MFCC mendapatkan hasil yang paling baik dibandingkan scenario lainnya karena hampir semua kelas sudah dapat diprediksi dengan benar. Namun masih terdapat kelas yang masih sering salah prediksi yaitu kelas netral dengan hanya memiliki 3 data yang ditebak dengan benar dari keseluruhan 7 data yang ada dikarenakan masih banyak data tertebak ke kelas tenang. Sedangkan untuk kelas yang paling sering diprediksi dengan benar adalah kelas tenang dengan banyak data yang ditebak dengan benar sebanyak 15 dari keseluruhan 19 data yang ada.



Gambar 9. *Confusion matrix* untuk pengujian menggunakan metode MFCC

Berdasarkan Tabel 8 untuk skema pengujian dengan menggunakan metode ekstraksi MFCC dihasilkan akurasi rata-rata untuk keseluruhan kelas emosi sebesar 70%. Serta nilai rata-rata masing-masing *f1-score*, *recall* dan presisi adalah 67%, 67%, dan 59%.

Tabel 8. Perbandingan nilai akurasi, *f1-score,* presisi dan *recall* dengan menggunakan MFCC

Kelas Emosi	F1-score (%)	Presisi (%)	Recall (%)	Rata-rata Akurasi (%)
Jijik	62	67	57	70
Marah	83	85	81	
Netral	33	33	33	
Sedih	77	92	67	
Senang	68	62	75	
Takut	63	69	58	
Tenang	71	65	79	
Terkejut	78	74	83	

Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11
ISSN 2502-3357 (*Online*)
ISSN 2503-0477 (*Print*)

4. Kesimpulan

Setelah dilakukannya beberapa skenario pengujian dapat disimpulkan bahwa pembagian dataset dengan ukuran 80% untuk *data training*, 10% untuk *data validation* dan 10% untuk *data testing* memberikan hasil paling baik untuk pembuatan model CNN. Serta hasil pengujian yang dilakukan memakai metode ekstraksi fitur *Mel-Frequency Cepstral Coefficients (MFCC)* mendapatkan akurasi tertinggi sebesar 70% diikuti dengan rata-rata *recall* dan presisi masing-masing 67% dan 68%. Untuk kelas emosi yang paling sering ditebak dengan benar adalah emosi marah, terkejut, sedih dan tenang dengan rata-rata prediksi benar sebesar 77%.

Saran yang diberikan penulis untuk penelitian selanjutnnya berdasarkan penelitian ini yaitu menambahkan *data input* berbahasa Indonesia sehingga model dapat lebih mudah untuk memahami jenis emosi pada suara yang berbahasa Indonesia.

5. Kontribusi Penulis

F. J. Tanudjaja: *Data collection, Formal Analysis, Investigation, Methodology, Visualization, dan Writing – original draft*. **E. Y. Puspaningrum:** *Supervision, Validation, dan review*. **Y. V. Via:** *Supervision, Validation, dan review*.

6. Declaration of Competing Interest

Penulis menyatakan tidak ada konflik kepentingan.

7. Referensi

Aini, Y. K., Santoso, T. B., & Dutono, T. (2021, Mei). Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia. Jurnal Komputer Terapan, 7(1), 143 - 152. doi:10.35143/jkt

Alghifari, M. F., Gunawan, T. S., & Kartiwi, M. (2018). Speech Emotion Recognition Using Deep Feedforward Neural Network. Indonesian Journal of Electrical Engineering and Computer Science, 10(2), 554-561. doi:DOI: 10.11591/ijeecs.v10.i2.pp554-561

Anggraini, N. A., & Fadillah, N. (2019). Analisis Deteksi Emosi Manusia dari Suara Percakapan Menggunakan Matlab dengan Metode KNN. InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan, 3(2), 280-283.

Anjaini, A., Gautama, A., & Anggis, N. (2019). Implementasi dan Analisis Simulasi Deteksi Emosi Melalui Pengenalan Suara Menggunakan Mel-Frequency Cepstrum Coefficient dan Hidden Markov Model Berbasis IOT. e-Proceeding of Engineering, 2100-2107.

Aouani, H., & Ayed, Y. B. (2018). Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder. 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1-5.

Berhil, S., Benlahma, H., & N. L. (2019). A review paper on artificial intelligence at the service of human resources management. The Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 18(1), 32-40.

Helmiyah, S., Fadlil, A., & Yudhana, A. (2018). EKSTRAKSI CIRI EMOSI MANUSIA BERDASARKAN UCAPAN MENGGUNAKAN MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC). Prosiding Seminar Nasional Sains dan Teknologi, 1(1), 93-98.

Jain, U., Nathani, K., Ruban, N., Raj, A. N., Zhuang, Z., & Mahesh, V. G. (2018). Cubic SVM Classifier Based Feature Extraction and Emotion Detection from Speech Signals. 2018 International Conference on Sensor Networks and Signal Processing (SNSP). doi:10.1109/SNSP.2018.00081

Kasyadi, F., Ilyas, R., & Annisa, N. M. (2021, Desember). Peningkatan Kemampuan Pengenalan Emosi melalui Suara dalam Bahasa Indonesia. Multimedia Artificial Intelligent Networking Database (MIND) Journal, 6(2), 194 - 204. doi:https://doi.org/10.26760/mindjournal.v6i2.194-204

Kasyidi, F., & Lestari, D. P. (2018). Identification of four class emotion from Indonesian spoken language using acoustic and lexical features . Journal of Physics: Conference Series, 971(012048).

Lasiman, J. J., & Lestari, D. P. (2018). Speech Emotion Recognition for Indonesian Language Using Long Short-Term Memory. 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 40-43. doi:10.1109/IC3INA.2018.8629525

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLOS ONE, 13(5).

Qayyum, A. B., Arefeen, A., & Shahnaz, C. (2019). Convolutional Neural Network (CNN) Based Speech-Emotion Recognition. 2019 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON), 122-125.

Widodo, Y. F., Sunardi, & Fadil, A. (2019). Identifikasi Suara Pada Sistem Presensi Karyawan Dengan Metode Ekstraksi MFCC. J-SAKTI (Jurnal Sains Komputer & Informatika), 3(1), 115-125.

ISSN 2502-3357 (*Online*)
Tanudjaja. F. J dkk./TEKNOLOGI 13 (2) 2023 1-11 ISSN 2503-0477 (*Print*)
Winursito, A., Hidayat, R., & Bejo, A. (2018). Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition. 2018 International Conference on Information and Communications Technology (ICOIACT, 379-383.

Yao, Z., Wang, Z., Liu, W., Liu, Y., & Pan, J. (2020). Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN, MS-CNN and LLD-RNN. Speech Communication, 120, 11-19.

Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical Signal Processing and Control, 47, 312-323.