PENGENALAN DIALEK DI SUMATERA SELATAN MENGGUNAKAN ALGORITMA *DEEP NEURAL NETWORK*

M. Rizki Putra[1], Bhakti Yudho Suprapto[1] dan Suci Dwijayanti[1]

[1]Teknik Elektro, Universitas Sriwijaya, Palembang
Corresponding author: sucidwijayanti@ft.unsri.ac.id

ABSTRAK: Suatu bahasa memiliki beragam dialek di setiap daerah. Hal ini dapat mempengaruhi perkembangan teknologi, khususnya dalam pengenalan suara. Namun, penelitian yang membahas tentang dialek Sumatera Selatan belum ada sehingga pada penelitian ini dikembangkan sistem pengenalan dialek daerah dari Sumatera Selatan dengan menggunakan model *deep neural network* (DNN). *Dataset* yang digunakan dalam penelitian ini adalah data primer dari 5 reponden yang merupakan penutur asli dari dialek yang digunakan, yang terdiri dari dialek Beliti, dialek Sekayu, dialek Palembang, dialek Lahat, dialek Muara Enim, dan bahasa Indonesia baku. Ciri-ciri sinyal suara yang diperoleh dari dataset adalah mel spectrogram, *short time fourier transform* (STFT), dan *mel frequency cepstral coefficient* (MFCC). Hasil pengujian menunjukkan bahwa model DNN yang menggunakan *optimizer* Adam dan *loss cross entropy* memiliki hasil yang cukup baik dengan input berupa ekstraksi mel spectrogram dan STFT. Akurasi tertinggi dicapai dalam mengenali dialek Beliti, yaitu 72,7% dan dialek Palembang 71,4 % jika ekstrasi ciri yang digunakan adalah mel spectrogram. Sedangkan untuk Bahasa Indonesia, akurasi tertinggi adalah dengan menggunakan ekstraksi ciri STFT, yaitu 71,4%. Model yang menggunakan ciri MFCC menunjukkan performansi yang paling rendah. Hasil ini menunjukkan bahwa mel spectrogram dan STFT dapat digunakan sebagai input DNN untuk pengenalan dialek. Hasil penelitian juga menunjukkan bahwa model DNN dapat memprediksi semua dialek, kecuali dialek Muara Enim. Hal ini dikarenakan dialek Muara Enim direkam pada ruang terbuka sehingga background noise mempengaruhi pengenalan dialek.

Kata Kunci: DNN, pengenalan dialek, mel spectrogram, STFT, MFCC

ABSTRACT: A language has various dialects in each region. It can affect the development of technology, especially in speech recognition. However, the South Sumatran dialects have not been discussed yet. Thus, this study developed a method to recognize dialects using the deep neural network (DNN) model. The dataset used in this study was primary data from 5 respondents who are native speakers of the dialect used, which consists of Beliti dialect, Sekayu dialect, Palembang dialect, Lahat dialect, Muara Enim dialect, and standard Indonesian. The characteristics of the voice signal obtained from the dataset are mel spectrogram, short-time Fourier transform (STFT), and mel frequency cepstral coefficient (MFCC). The test results showed that the DNN model that uses the Adam optimizer and loss cross-entropy has good accuracy value with input in the form of mel spectrogram and STFT. The highest accuracy was achieved in recognizing the Beliti dialect at 73% and the Palembang dialect at 71% when using mel spectrogram features. As for Indonesian, the highest accuracy was achieved using STFT feature extraction, which was 71%. Meanwhile, MFCC showed the lowest performance. These results indicate that the mel spectrogram and STFT can be used as DNN input for dialect recognition. The results also showed that the DNN model can predict all dialects, except the Muara Enim dialect. This is because the Muara Enim dialect was recorded in an open space so that background noise affected dialect recognition

Keywords: DNN, dialect recognition, mel spectrogram, STFT, MFCC

PENDAHULUAN

Dialek merupakan variasi bahasa yang merupakan karakteristik kelompok tertentu pada penutur suatu bahasa (Lexico, 2020). Jenis dialek terbagi menjadi empat, yaitu: dialek regional, dialek etnis, *sociolect*, dan aksen. Dialek regional merupakan variasi bahasa yang dipengaruhi oleh kondisi geografis suatu wilayah. Dialek etnis merupakan

variasi bahasa yang berhubungan dengan kelompok etnis tertentu. *Sociolect* merupakan jenis dialek yang dipengaruhi oleh kelompok-kelompok sosial. Sedangkan, aksen adalah hasil dari pembedaan fonetik atau pengucapan antara kelompok satu dengan yang lain (JTA Technology Consulting, 2020).

Salah satu cara untuk mengetahui asal usul seorang penutur adalah dengan mendengar cara/gaya berbicaranya. Setiap penutur memiliki dialek tersendiri yang didapatkan dari keluarga, kerabat dan lingkungan sekitarnya. Terkadang, dialek kedua orang dalam sebuah dialog dapat terdengar sama meskipun memiliki latar belakang daerah yang berbeda ketika berbicara. Selain itu, ada kemungkinan seseorang yang berasal dari daerah lain berbicara dengan dialek yang berbeda dengan daerah asalnya. Dialek akan mempengaruhi pengenalan suara dan berbagai aplikasi suara lainnya.

Penelitian dialek untuk bahasa Indonesia juga sangat penting karena Indonesia sendiri memiliki lebih dari 700 bahasa yang dituturkan (Lewis, 2008). Keberagaman etnis dan bahasa daerah di indonesia merupakan salah satu faktor terbentuknya keberagaman dialek pada bahasa indonesia. Kondisi geografis indonesia juga memberi keunikan dialek dimana terdapat banyak bahasa daerah yang berbeda satu sama lain, baik pada pulau yang sama atau antar pulau.

Salah satu penelitian yang membahas pengenalan dialek di Indonesia yaitu dialek Jawa dan Sunda dilakukan oleh R. Rahmawati dan D. P. Lestari. Pada penelitian tersebut digunakan metode *Gaussian Mixture Model* (GMM) dan *I-Vector* (Rahmawati et al. 2018). Namun, hasil yang diperoleh belum akurat karena GMM sangat bergantung pada ekstrasi ciri yang digunakan.

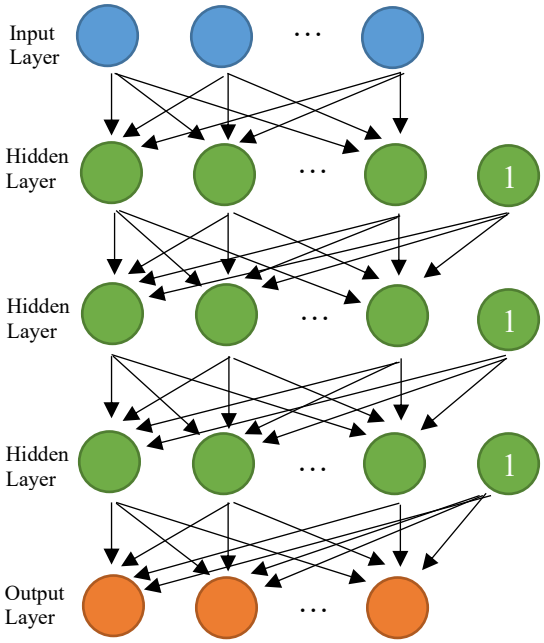
Lalu, penelitian yang telah dilakukan oleh Warrohmah *et al*. (2018) mengindentifikasi dialek penutur pada bahasa Indonesia dimana penutur memiliki dialek Jawa, Batak, dan Minang. Metode yang digunakan adalah *back propagation neural network* (BPNN) dengan ekstrasi ciri berupa *mel frequency cepstal coefficient* (MFCC). Ada pula penelitian mengenai pengenalan dialek Garut pada bahasa Sunda Selatan dengan memakai metode *recurrent neural network* (RNN) dan ekstraksi ciri yang sama (Hakim et al. 2018).

Namun, penelitian yang membahas tentang dialek Sumatera Selatan belum ditemukan. Padahal, Sumatera Selatan memiliki keanekaragaman dialek yang dituturkan, diantaranya adalah penutur bahasa Melayu Palembang, bahasa Ogan, bahasa Komering, bahasa Semendo, bahasa Lahat, dan bahasa-bahasa lain di Sumatera Selatan. Terlebih lagi, Indonesia memiliki 718 bahasa daerah berdasarkan data dari Dapobas (Kementerian Pendidikan dan Kebudayaan, 2020). Sehingga, pada penelitian ini akan dikembangkan sistem identifikasi dialek yang ada di Sumatera Selatan. Berbeda dengan penelitian-penelitian sebelumnya yang sangat bergantung pada ekstrasi ciri, penelitian ini menggunakan metode *deep neural network* (DNN).

Penelitian berkaiatan dengan pengenalan dialek menggunakan DNN telah dilakukan oleh Ruan *et al* (2017) untuk pengenalan dialek Lhasa pada bahasa Tibet. Fitur yang diekstrak dengan metode *deep learning* memiliki kemampuan pemodelan yang lebih baik daripada sinyal-sinyal model *hidden markov model* (HMM) fonem tradisional, sehingga performa model pengenalan suara telah meningkat. Penelitian tersebut menunjukkan metode yang lebih baik dalam menggantikan GMM dengan DNN untuk pengenalan dialek bahasa Tibet Lhasa. Sehingga, pada penelitian ini akan digunakan DNN untuk mengenali dialek yang ada di Sumatera Selatan. Selain itu, data suara yang digunakan berasal langsung dari subjek yang berasal dari daerah asal penutur bahasa asli.

METODE PENELITIAN

Deep Neural Network



Gambar 1 Arsitektur *Deep Neural Network* (DNN)

Deep neural network (DNN) merupakan *multilayer perceptron* (MLP) konvensional dengan banyak (lebih dari dua) *hidden layer* (Yu dan Deng, 2020). Gambar 1 menunjukkan sebuah DNN dengan total lima *layer* yang mencakup satu *input layer*, tiga *hidden layer*, dan satu *output layer*. Untuk penyederhanaan notasi, disini

ditunjukkan lapisan *input* sebagai layer 0 dan lapisan *output* sebagai *layer L* untuk layer yang berjumlah $L + 1$-*layer*.

Pada *layer* L pertama,

$$\mathbf{v}^{\boldsymbol{\ell}} = \boldsymbol{f}\left(\mathbf{z}^{\boldsymbol{\ell}}\right) = \boldsymbol{f}\left(\mathbf{W}^{\boldsymbol{\ell}}\mathbf{v}^{\boldsymbol{\ell-1}} + \mathbf{b}^{\boldsymbol{\ell}}\right), \text{ untuk } \mathbf{0} < \boldsymbol{\ell} < \boldsymbol{L}, \quad (1)$$

dimana $\mathbf{z}^{\ell} = \mathbf{W}^{\ell}\mathbf{v}^{\ell-1} + \mathbf{b}^{\ell} \in \mathbb{R}^{N_\ell \times 1}$, $\mathbf{v}^{\ell} \in \mathbb{R}^{N_\ell \times 1}$, $\mathbf{W}^{\ell} \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$, $\mathbf{b}^{\ell} \in \mathbb{R}^{N_\ell \times 1}$ dan $N_\ell \in \mathbb{R}$ berturut-turut adalah vektor eksitasi, vektor aktivasi, matriks bobot, vektor bias, dan jumlah neuron pada *layer* ℓ. $\mathbf{v}^0 = \mathbf{o} \in \mathbb{R}^{N_0 \times 1}$ adalah vektor observasi (atau fitur), $N_0 = D$ adalah dimensi fitur, dan $f(\cdot) : \mathbb{R}^{N_\ell \times 1} \rightarrow \mathbb{R}^{N_\ell \times 1}$ adalah fungsi aktivasi yang diterapkan kepada *element-wise* vektor eksitasi.

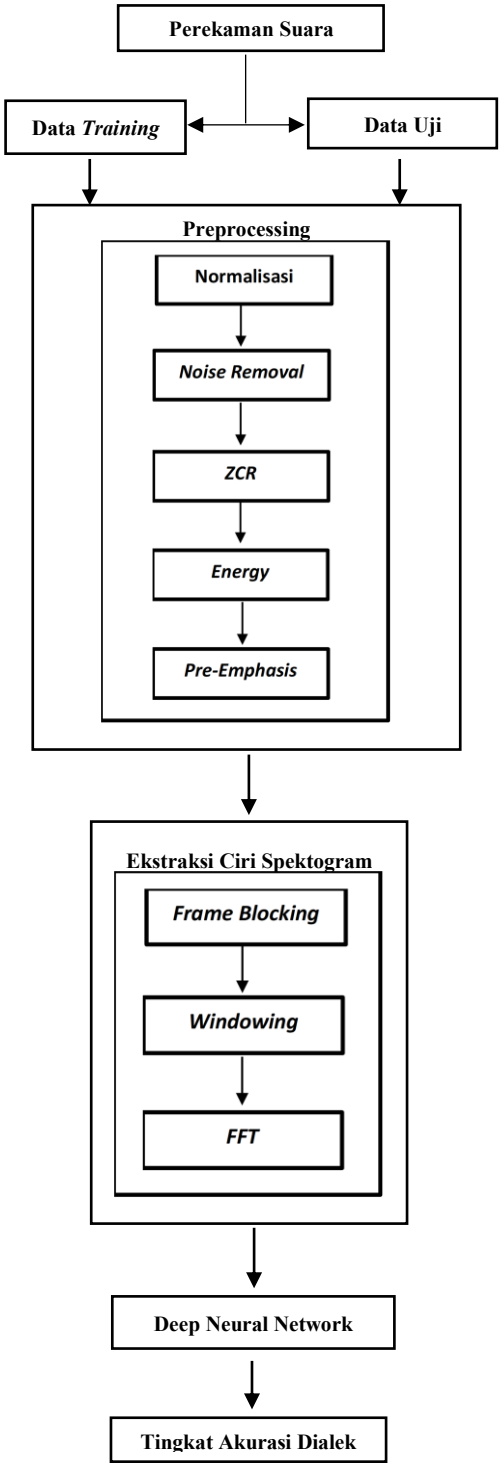
Sistematika Perancangan Penelitian

Penelitian ini dilakukan untuk mengidentifikasi logat pada penutur bahasa Indonesia dari daerah Sumatera Selatan. Objek penelitian yang akan digunakan adalah sinyal suara dari mahasiswa jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya yang merupakan penutur asli dari beberapa daerah yang berada di Sumatera Selatan.

Data primer yang digunakan berasal langsung dari penutur asli dari dialek yang akan diuji, yaitu dialek Palembang, dialek Muara Enim, dialek Lahat, dialek Sekayu dan dialek Musi Rawas. Penutur yang akan diambil sampel suaranya berjumlah delapan orang, yang terdiri dari laki-laki dan perempuan mewakili satu dialek. Perekaman akan dilakukan dengan menggunakan mikrofon, dimana penutur akan mengucapkan masing-masing lima kalimat dengan lima kali pengulangan. Sampling frekuensi yang akan digunakan adalah 16 kHz. Sementara itu, informasi suara akan diambil di ruangan tertutup. Para penutur akan membaca teks dalam dua tahap, yaitu membaca sesuai dengan teks (bahasa Indonesia baku) dan menerjemahkannya ke dalam dialek daerah.

Penelitian ini menggunakan metode DNN yang akan diprogram dengan menggunakan perangkat lunak Python. Alur proses pengenalan dapat dilihat pada alur proses penelitian yang ada pada Gambar 2. Pertama, data suara akan diambil dengan merekam suara penutur. Setelah data suara diperoleh, jumlah data yang diperoleh tersebut akan dibagi menjadi data *training* dan data uji. Kemudian, suara yang telah diperoleh akan memasuki tahap *pre-processing* dimana derau pada sampel suara akan dihilangkan. Setelah itu, ciri-ciri spektogram pada sampel suara akan diperoleh dari ekstraksi ciri. Jenis ciri yang digunakan penelitian ini diantaranya adalah mel spectrogram, *short time Fourier Transform* (STFT), dan MFCC.

Ciri-ciri yang telah diperoleh dilatih menggunakan algoritma DNN untuk klasifikasi dialek. Setelah proses pelatihan, dilakukanlah proses pengujian, dimana sistem pengenalan dialek akan mengenali data uji. Output pada sistem ini berupa tingkat akurasi dari klasifikasi dialek.



Gambar 2 Alur Proses Penelitian

Pengujian model DNN pengenalan dialek dilakukan untuk mengetahui kinerja sistem dalam

mengenali dialek yang dibawa oleh penutur suara. Persentase tersebut akan dimuat ke dalam *confusion matrix*. Dengan *confusion matrix*, kinerja sistem pengenalan dialek yang dirancang dapat diketahui. Selain itu, nilai akurasi, *recall* dan presisi pada masing-masing dialek juga akan dihitung. Nilai *recall* dan presisi dihitung dengan menggunakan persamaan berikut:

$$\boldsymbol{Akurasi} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \quad (2)$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} \quad (3)$$

$$\textit{Presisi} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}} \quad (4)$$

dimana TP merupakan nilai *true positive*, FP merupakan nilai *false positive*, TN adalah nilai *true negative*, dan FN merupakan nilai *false negative*.

Setelah nilai *recall* dan presisi telah diperoleh, nilai skor F1 pada masing-masing dialek akan dihitung dengan menggunakan persamaan berikut:

$$\textit{F1-score} = \mathbf{2} \cdot \frac{\textit{Presisi} \cdot \boldsymbol{Recall}}{\textit{Presisi} + \boldsymbol{Recall}} \quad (5)$$

Nilai *F1-score* pada sistem akan diperoleh dengan menghitung rata-rata aritmatik dari nilai-nilai presisi dan *recall* yang telah diperoleh.

HASIL DAN PEMBAHASAN

Pengambilan Data Rekaman Suara

Data rekaman diperoleh menggunakan Microphone FIFINE K669B. Parameter yang digunakan dalam proses pengumpulan suara yaitu *mono-channel* dengan *sample rate* sebesar 16.000 Hz.

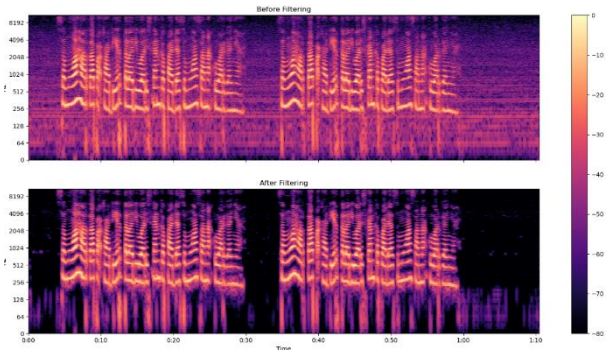
Data merupakan data primer yang diperoleh dari enam narasumber yang membaca teks dalam bahasa Indonesia dan logat daerah asal. Setiap narasumber membaca dua puluh teks. Setiap teks akan diulang pembacaannya sebanyak dua kali dalam satu perekaman.

Data rekaman suara yang didapatkan berjumlah 200 data yang terdiri dari 100 audio Bahasa Indonesia, 20 audio dialek Sekayu, 20 audio dialek Beliti, 20 audio dialek Palembang, 20 audio dialek Lahat, dan 20 audio dialek Muara Enim. Kemudian, *dataset* tersebut dibagi menjadi tiga kelompok: data *training*, data validasi, dan data uji dengan proporsi 3:1:1.

Proses *Preprocessing* Sinyal Suara

Setelah sinyal suara didapatkan, sinyal suara tersebut dipraposes. *Preprocessing* diterapkan untuk memperbaiki sinyal suara yang masih mengandung *noise* yang tidak diinginkan. Hal ini bertujuan agar *noise* yang terdapat pada fitur tidak memperburuk kinerja model pengenalan dialek.

Proses ini didahului dengan pemfilteran sinyal suara. Sinyal suara yang diperoleh akan dikurangi deraunya dengan menerapkan algoritma reduksi derau (*noise reduction*). Algoritma ini berdasarkan pada metode *spectral gating* yang merupakan salah satu bentuk *noise* gate. Algoritma tersebut bekerja dengan mengkomputasi spektogram sinyal dan memperkirakan ambang batas derau bagi tiap-tiap pita frekuensi dari sinyal/derau. Ambang batas itu digunakan untuk mengkomputasi *mask* yang menghalangi derau di bawah ambang batas frekuensi (Timsainb and Sainburg, 2020). Perbandingan spectrogram suara sebelum dan sesudah *filtering* dapat dilihat pada Gambar 3.



Gambar 3 Perbandingan *Spectrogram* Sinyal Suara Sebelum *Filtering* dan Setelah *Filtering*

Ciri pada *dataset* audio yang telah diolah pada tahap *preprocessing* akan diekstrak menjadi beberapa ciri. Hal ini dilakukan dengan menggunakan *library* Librosa. Jenis ciri yang diuji pada penelitian ini diantaranya adalah Mel spectrogram, STFT, dan MFCC. Dalam ekstraksi ketiga ciri, parameter n-FFT yang digunakan adalah sebesar 2048, *hop length* sebesar 64, dan semua *windowing* menggunakan *window* Hann. Algoritma ekstraksi ciri sinyal suara akan menghasilkan 265 input mel, 1025 input STFT, dan 20 input MFCC. Masing-masing jenis ciri sinyal suara disimpan ke dalam file .csv lalu dilanjutkan ke proses pelatihan data.

Proses *Training Dataset*

Kumpulan ciri sinyal suara akan dilatih menggunakan model *deep neural network* (DNN) yang dirancang untuk klasifikasi data. Model ini disusun dengan lima *hidden layer* dalam 5000 *epoch*. *Hidden layer* ini merupakan parameter terbaik setelah dilakukan pelatihan dengan

menggunakan jumlah *hidden layer* yang berbeda. Data validasi dan data uji telah ditentukan dengan proporsi 1:1.

Pelatihan DNN ini dilakukan dengan menggunakan *input*, yaitu *mel spectrogram*, STFT, dan MFCC yang sebelumnya telah diubah menjadi bentuk vektor. Adapun parameter yang digunakan pada pelatihan dengan menggunakan masing-masing input dapat dilihat pada Tabel 1.

Tabel 1 Parameter Pelatihan DNN

Parameter	**Nilai Parameter**
Optimizer	Adam
	Stochastic Gradient Descent (SGD)
Loss function	MSE
	Categorical cross entropy
Learning rate	0,01 (SGD)
	0,001 (Adam)

Setiap input ciri dari sinyal suara yang didapat akan dilatih dengan menggunakan arsitektur DNN yang sama. Perbedaan pada pelatihan ketiga ciri tersebut adalah pada *input.* Arsitektur DNN yang digunakan pada proses pelatihan ciri mel spectrogram dapat dilihat pada Tabel 2.

Tabel 2 Arsitektur Model DNN untuk Pelatihan Ciri Mel Spectrogram

Layer	**Input**	**Fungsi Aktivasi**	***Dropout***
1	256	ReLU	0.6
2	128	ReLU	0.6
3	128	ReLU	0.5
4	128	ReLU	0.5
5	128	ReLU	0.5
6	128	ReLU	0.5
7	6	Softmax	

Arsitektur DNN pada proses pelatihan ciri STFT adalah seperti yang ditunjukkan pada Tabel 3.

Tabel 3 Arsitektur Model DNN untuk Pelatihan Ciri STFT

Layer	**Input**	**Fungsi Aktivasi**	***Dropout***
1	1025	ReLU	0.6
2	128	ReLU	0.6
3	128	ReLU	0.5
4	128	ReLU	0.5
5	128	ReLU	0.5
6	128	ReLU	0.5
7	6	Softmax	

Untuk pelatihan ciri MFCC disesuaikan input fiturnya, dimana input pada *input layer* menjadi 20 dan pada *neuron* pada *hidden layer* menjadi 10, seperti yang dapat dilihat pada Tabel 4.

Tabel 4 Arsitektur Model DNN untuk Pelatihan Ciri MFCC

Layer	**Input**	**Fungsi Aktivasi**	***Dropout***
1	20	ReLU	0.6
2	10	ReLU	0.6
3	10	ReLU	0.5
4	10	ReLU	0.5
5	10	ReLU	0.5
6	10	ReLU	0.5
7	6	Softmax	

Pelatihan Dataset Ciri

Mel Spectrogram

Perbandingan kinerja pelatihan DNN untuk beberapa parameter *optimizer* dan *loss function* dengan menggunakan fitur *mel spectrogram* dapat dilihat pada Tabel 5. Pelatihan model *mel spectrogram* memperoleh hasil akhir dengan nilai *training loss* terkecil berada pada model dengan *optimizer* Adam. Sementara itu, nilai *loss* validasi terkecil diperoleh pada model dengan *optimizer* SGD. Nilai akurasi terbesar juga diperoleh pada model dengan *optimizer* SGD. Untuk model dengan *loss categorical cross entropy*, diperoleh nilai *loss*, *validation loss*, dan akurasi yang kecil dibandingkan dengan *loss* MSE. Hasil ini menunjukkan bahwa model DNN yang menggunakan *loss function* MSE dan *optimizer* SGD memiliki kemampuan untuk melakukan generalisasi yang lebih baik dibandingkan dengan menggunakan kombinasi SGD dan *cross entropy* pada saat input DNN adalah *mel spectrogram*.

Tabel 5 Perbandingan Kinerja Pelatihan Model DNN Menggunakan Fitur Mel Spectrogram

	SGD-MSE	**Adam-MSE**	**SGD - *Cross entropy***	**Adam-*Cross entropy***
loss	0.0904	0.0471	0.6665	0.4560
val_loss	0.1196	0.1622	17.9482	95.3096
val_accuracy	0.5000	0.4500	0.3500	0.3500

Short-Time Fourier Transform (STFT)

Pada Tabel 6 terlihat bahwa nilai *training loss* pada model yang menggunakan fungsi loss MSE cenderung lebih tinggi dibandingkan dengan model yang

menggunakan fungsi *loss categorical cross entropy*. Selain itu, model yang menggunakan fungsi *loss categorical cross entropy* memperoleh loss validasi yang lebih tinggi dibandingkan dengan model yang menggunakan fungsi *loss* MSE. Selain itu, model DNN yang menggunakan *optimizer* Adam memberikan nilai *training loss* yang lebih kecil namun *validation loss* yang lebih besar dari model yang menggunakan *optimizer* SGD. Nilai akurasi validasi terbesar diperoleh pada model SGD-MSE dan yang terkecil diperoleh pada model Adam-MSE. Hasil ini menunjukkan bahwa model DNN yang menggunakan fungsi *loss* MSE dan *optimizer* SGD memiliki kemampuan untuk melakukan generalisasi yang lebih baik dibandingkan dengan menggunakan kombinasi Adam dan MSE Ketika input yang digunakan adalah STFT**.**

Tabel 6 Perbandingan Kinerja Pelatihan Model DNN Menggunakan Fitur STFT

	SGD-MSE	**Adam-MSE**	**SGD - *Cross entropy***	**Adam-*Cross entropy***
loss	0.0706	0.0678	0.4485	0.3379
val_loss	0.1204	0.1675	33.8402	72.3370
val_accuracy	0.5000	0.3750	0.4750	0.4000

Mel-Frequency Cepstrum Coefficient (MFCC)

Perbandingan kinerja pelatihan untuk berbagai *loss function* dan *optimizer* dapat dilihat pada Tabel 7. Berdasarkan tabel dapat dilihat bahwa nilai *loss* pada model yang menggunakan fungsi *loss categorical cross entropy* lebih tinggi dibandingkan dengan model yang menggunakan fungsi *loss* MSE. Lalu, nilai *validation loss* pada model yang menggunakan fungsi loss *categorical cross entropy* lebih tinggi dibandingkan dengan model yang menggunakan fungsi *loss* MSE. Nilai akurasi validasi yang diperoleh pada model yang menggunakan fungsi *loss* MSE lebih tinggi dari model yang menggunakan fungsi *loss categorical crossentropy*.

Sementara itu, nilai *training loss* pada model DNN yang menggunakan *optimizer* SGD lebih kecil dibandingkan dengan model yang menggunakan *optimizer* Adam. Kemudian, nilai *validation loss* pada model yang menggunakan *optimizer* Adam lebih tinggi dari model yang menggunakan model yang menggunakan *optimizer* SGD. Nilai akurasi validasi yang diperoleh pada model yang menggunakan *optimizer* SGD lebih tinggi dibandingkan dengan model yang menggunakan *optimizer* Adam. Hasil ini menunjukkan bahwa model DNN yang menggunakan fungsi *loss* MSE dan *optimizer* SGD memiliki kemampuan untuk melakukan generalisasi yang lebih baik dibandingkan dengan menggunakan kombinasi Adam dan *categorical cross entropy* pada saat input berupa MFCC.

Tabel 7 Perbandingan Kinerja Pelatihan Model DNN Menggunakan Fitur MFCC

	SGD-MSE	**Adam-MSE**	**SGD - *Cross entropy***	**Adam-*Cross entropy***
loss	0.1165	0.0927	1.1306	1.1713
val_loss	0.1081	0.1725	7.3791	106.2369
val_accuracy	0.5750	0.4750	0.5000	0.2250

Berdasarkan Tabel 5, 6, dan 7 dapat dilihat bahwa nilai *loss* terendah didapatkan ketika input yang digunakan adalah STFT dan *mel spectrogram.* Sedangkan *loss* tertinggi didapatkan untuk input MFCC. Hasil ini menunjukkan bahwa STFT dan *mel spectrogram* memiliki informasi sinyal suara yang lebih baik dibandingkan MFCC. Selain itu, penggunaan ekstraksi ciri STFT dan *mel spectrogram* dapat menghindari *cost computation* karena perhitungan MFCC yang kompleks.

Pengujian *Dataset*

Proses pengujian *dataset* dilakukan dengan memprediksi label *dataset* uji yang sebelumnya dipilih secara acak dengan *randomizer*. Hasil prediksi model DNN nantinya akan menunjukkan nilai akurasi, *recall*, presisi, dan F1-*score*.

Nilai F1-score hasil prediksi dengan menggunakan *optimizer* SGD dan fungsi *loss* MSE dapat dilihat pada Tabel 8. Dari tabel tersebut, dapat dilihat bahwa F1-*score* untuk ciri *mel spectrogram* dan STFT lebih baik dibandingkan dengan MFCC dengan nilai sebesar 0,667 untuk Bahasa Indonesia Hasil ini menunjukkan bahwa model DNN belum cukup baik digunakan untuk memprediksi dialek spesifik dari Sumatera Selatan.

Pada dialek bahasa Indonesia baku, nilai F1-*score*, pada model yang menggunakan ciri-ciri MFCC memperoleh nilai yang lebih rendah daripada model yang menggunakan ciri-ciri yang lain. Hal ini disebabkan karena ada informasi yang hilang pada saat proses MFCC dibandingkan dengan STFT dan *mel spectrogram*.

Tabel 9 menunjukkan F1-score hasil prediksi dengan menggunakan *optimizer* Adam dan fungsi *loss* MSE. Berdasarkan hasil prediksi pada tabel tersebut, model yang mengolah ciri-ciri *mel pectrogram* dapat mengenali sebagian dialek pada *dataset* selain dialek Muara Enim dan dialek Sekayu. Selain itu, model tersebut dapat mengenali dataset dialek Palembang dengan baik, dimana nilai nilai F1-*score* yang didapat adalah 0,727. Hal ini dapat disebabkan karena Bahasa Palembang memiliki

kecenderungan yang hampir sama dengan Bahasa Indonesia dan perbedaan utama dapat dilihat pada penggunaan vokal /o/ yang dominan.

Tabel 8 Nilai F1-*Score* pada Prediksi Model DNN Menggunakan Optimizer SGD dan Fungsi *Loss* MSE

Dialek	Mel Spectrogram	STFT	MFCC
BLT	0	0	0
IDN	0.667	0.667	0.525
LHT	0	0	0
MEN	0	0	0
PLG	0	0	0
SKY	0	0	0

Ketika menggunakan ekstraksi ciri STFT, hasil prediksi menunjukkan bahwa model tersebut memiliki performansi terbaik pada saat mengenali dialek bahasa Indonesia baku dengan nilai F1-score 0.653. Sementara itu, model yang mengolah ciri-ciri MFCC tidak dapat mengenali dialek daerah.

Tabel 9 Nilai F1-*Score* pada Prediksi Model DNN Menggunakan *Optimizer* Adam dan Fungsi *Loss* MSE

Dialek	Mel Spectrogram	STFT	MFCC
BLT	0.5	0.286	0
IDN	0.619	0.653	0.686
LHT	0.167	0.286	0
MEN	0	0.2	0
PLG	0.727	0	0
SKY	0	0	0

F1-*score* hasil prediksi dengan menggunakan *optimizer* SGD dan fungsi loss *categorical cross entropy* dapat dilihat pada Tabel 10. Berdasarkan tabel tersebut, model DNN yang menggunakan Mel Spectrogram hanya mampu mengenali dialek Beliti dan dialek bahasa Indonesia baku. Untuk performa pengenalan dialek yang terbaik pada model ini diperoleh kepada dialek bahasa Indonesia baku dengan nilai F1-score 0,622.

Pada model DNN yang menggunakan ciri-ciri STFT, hanya dialek Beliti, dialek bahasa Indonesia baku, dan dialek Sekayu yang dapat dikenali. Dialek dengan performa pengenalan terbaik diperoleh kepada dialek bahasa Indonesia baku dengan nilai F1-score 0,625. Sementara itu, model yang mengolah ciri-ciri MFCC hanya mampu mengenali dialek bahasa Indonesia.

Tabel 10 Nilai F1-*Score* pada Prediksi Model DNN Menggunakan Optimizer SGD dan Fungsi *Loss Categorical Cross entropy*

Dialek	Mel Spectrogram	STFT	MFCC
BLT	0.571	0.222	0
IDN	0.622	0.625	0.667
LHT	0	0	0
MEN	0	0	0
PLG	0	0	0
SKY	0	0.333	0

Tabel 11 menunjukkan F1-*score* hasil prediksi dengan menggunakan *optimizer* Adam dan fungsi *loss categorical cross entropy*. Ketika menggunakan input *mel spectrogram*, performa pengenalan dialek yang terbaik dalam model ini diperoleh pada dialek Beliti dengan F1-score 0,727 (72,7%), yang kemudian dialek Palembang dengan F1-score 0.714 (71,4%), dialek bahasa Indonesia baku dengan F1-score 0,684 (68,4%), dan dialek Sekayu dengan F1-score 0,182.

Kemudian pada model DNN yang menggunakan ciri-ciri STFT, hasil prediksi menunjukkan bahwa model ini dapat mengenali semua dialek selain dialek Muara Enim. Dialek dengan performa pengenalan yang terbaik pada model ini diperoleh kepada dialek bahasa Indonesia baku dengan F1-*score* 0,714 atau 71,4%. Sementara itu, model yang memproses ciri-ciri MFCC hanya mampu mengenali dialek Beliti, dialek bahasa Indonesia baku dan dialek Lahat. Model ini tidak dapat mengenali dialek Palembang dan dialek Sekayu sama sekali, serta salah dalam mengenali dialek Muara Enim. Dialek dengan performa pengenalan yang terbaik pada model ini diperoleh kepada dialek bahasa Indonesia baku dengan F1-score 0,64 atau 64%.

Tabel 11 Nilai F1-*Score* pada Prediksi Model DNN Menggunakan Optimizer Adam dan Fungsi *Loss Categorical Cross entropy*

Dialek	Mel Spectrogram	STFT	MFCC
BLT	0.727	0.444	0.222
IDN	0.684	0.714	0.64
LHT	0	0.4	0.333
MEN	0	0	0
PLG	0.714	0.4	0
SKY	0.182	0.4	0

Berdasarkan hasil pengujian yang dilakukan, model DNN yang cukup baik dan dapat memprediksi dialek yang ada di Sumatera Selatan adalah model DNN yang

menggunakan *optimizer* Adam dan *loss categorical cross entropy*. Hasil ini juga menunjukkan bahwa dialek di Sumatera Selatan yang dapat diprediksi dengan baik adalah dialek Beliti, Lahat, Palembang, dan Sekayu. Sedangkan dialek Muara Enim adalah dialek yang paling sulit untuk dikenali oleh semua model DNN, baik dengan menggunakan ekstrasi ciri *mel spectrogram* dan STFT. Hal ini dapat disebabkan oleh pengambilan data sampel narasumber yang dilakukan pada ruang terbuka dibandingkan dengan sampel dari narasumber yang lain. *Noise background* ini mempengaruhi hasil prediksi dari dialek Muara Enim. Selanjutnya, ekstraksi ciri yang baik digunakan dalam pengenalan dialek adalah *mel spectrogram* dan STFT.

KESIMPULAN DAN SARAN

Berdasarkan dari hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa model DNN yang terbaik dalam pengenalan dialek di Sumatera Selatan dalah model DNN yang dilatih dengan menggunakan optimizer Adam dan loss categorical cross entropy. Selanjutnya, ekstraksi ciri mel spectrogram dan STFT dapat digunakan sebagai input DNN dan ciri ini memerlukan proses komputasi yang tidak terlalu kompleks dibandingkan dengan ciri mel frequency cepstral coefficient (MFCC). Jika dilihat lebih detail untuk masing-masing dialek, model DNN yang menggunakan optimizer Adam mampu memprediksi dialek yang ada di Sumatera Selatan dengan baik, meliputi dialek Beliti, Lahat, Palembang, dan Sekayu. Akurasi tertinggi dicapai untuk dialek beliti yaitu 72,7% dan dialek Palembang 71,4 % jika ekstrasi ciri yang digunakan adalah mel spectrogram.Sedangkan untuk bahasa Indonesia, akurasi tertinggi adalah dengan menggunakan ekstraksi ciri STFT, yaitu 71,4%. Berdasarkan hasil penelitian, dialek Muara Enim merupakan dialek yang paling sulit untuk dikenali yang disebabkan oleh *background noise* yang mempengaruhi akurasi pengenalan dialek. Penelitian ini juga menunjukkan bahwa dialek bahasa Indonesia lebih mudah dikenali dibandingkan dengan dialek daerah lainnya.

Untuk penelitian selanjutnya, jumlah dialek yang digunakan perlu ditambah agar terdapat variasi yang lebih banyak pada proses pelatihan DNN. Selain itu, penelitian pengenalan dialek dapat mempertimbangkan arsitektur *deep learning* yang lain

DAFTAR PUSTAKA

Hakim, L. A., Osmond, A. B., and Saputra, R. E. (2018). Recurrent Neural Network Untuk Pengenalan Ucapan Pada Recurrent Neural Network for Speech Recognition.

JTA Technology Consulting. "English Structures: Sociolingustics" (pg. 2) [Online]. Available: http://web.mnstate.edu/houtsli/tesl551/Socio/page2.htm. [Accessed: Jan 6, 2021]

Kementerian Pendidikan dan Kebudayaan, "Bahasa Daerah di Indonesia", *Kementerian Pendidikan dan Kebudayaan.* [Online]. Available: https://dapobas.kemdikbud.go.id/homecat.php?show=url/petabahasa&cat=6/. [Accessed: Jul. 14, 2020].

Lewis, M. P. (2009). Ethnologue: Languages of the World (sixteenth ed.), SIL International.

Lexico, "Dialect." [Online]. Available: https://www.lexico.com/definition/dialect/. [Accessed: 14-Jul-2020].

Mawadda Warohma, A., Kurniasari, P., Dwijayanti S., Irmawan and Yudho Suprapto, B. (2018). "Identification of Regional Dialects Using Mel Frequency Cepstral Coefficients (MFCCs) and Neural Network," *2018 International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, pp. 522-527.

Rahmawati, R. and Lestari, D. P. (2018). "Java and Sunda dialect recognition from Indonesian speech using GMM and I-Vector", *Proceeding 2017 11th Int. Conf. Telecommun. Syst. Serv. Appl. TSSA 2017*, vol. 2018-Janua, pp. 1–5.

Ruan, W., Gan, Z., Liu, B., and Guo, Y. (2017). "An Improved Tibetan Lhasa Speech Recognition Method Based on Deep Neural Network," *Proc. - 10th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2017*, vol. 2017-Octob, pp. 303–306.

Timsainb and Sainburg, T., "noisereduce". [Online]. Available: https://github.com/timsainb/noisereduce. [Accessed Sept. 26, 2021].

Yu, D. and Deng, L. (2020). Automatic Speech Recognition: A Deep Learning Approach.