

IDENTIFIKASI ASAL DAERAH BERDASARKAN SUARA MANUSIA DENGAN METODE *LINIER PREDICTIVE CODING* (LPC)

IDENTIFICATION OF REGIONAL ORIGIN BASED ON HUMAN VOICE USING LINEAR PREDICTIVE CODING (LPC) METHOD

Musliha Syam¹, Dr. Ir. Jangkung Raharjo², Raditiana Patmasari, S.T.,M.T.³

^{1,2,3} Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

¹muslihasyam@student.telkomuniversity.ac.id, ²jangkungraharjo@telkomuniversity.co.id,

³raditianapatmasari@telkomuniversity.ac.id

Abstrak

Suara manusia memiliki ciri khas masing-masing. Pengucapan kata yang sama akan menghasilkan pola sinyal yang berbeda sesuai dengan aksent atau pelafalannya. Pada penelitian ini telah dibuat sistem yang mampu mengidentifikasi asal daerah berdasarkan suara manusia. Asal daerah yang akan diidentifikasi adalah Sulawesi Selatan, Sumatra Barat, Jawa Barat dan Jawa Tengah..

Pada penelitian ini telah dibuat sistem untuk mengidentifikasi asal daerah berdasarkan suara manusia dengan metode ekstraksi ciri Linier Predictive Coding (LPC) dan klasifikasi Jaringan Saraf Tiruan (JST) dengan metode Propagasi Balik. Mula-mula data suara disimpan dalam file wav, kemudian dianalisa dengan LPC. Koefisien LPC yang diperoleh dari analisa LPC selanjutnya menjadi vektor masukan JST untuk melatih jaringan. Hasil pelatihan berupa perubahan bobot jaringan sehingga diperoleh jaringan yang memiliki kemampuan pengklasifikasian. Jaringan tersebut kemudian diuji dengan mensimulasikannya menggunakan data latih dan data uji guna menghasilkan persentase keberhasilan pengenalan. Pengujian dilakukan dengan beberapa perubahan nilai parameter agar diperoleh persentase pengenalan tertinggi..

Hasil pengujian menunjukkan nilai akurasi tertinggi yang didapatkan yaitu 80.76% dengan menggunakan 52 data latih dan 52 data uji. Parameter terbaik yang digunakan yaitu 3 hidden layer pada JST.

Kata kunci : *Linier Predictive Coding (LPC), Jaringan Saraf Tiruan (JST), Suara, Aksent.*

Abstract

Each human voice has the characteristics of themselves. Pronouncing the same word will produce a different signal pattern according to the accent or pronunciation. In this research has been made a system that can identify the place of origin based on human voice. The places of origin that has been identified are South Sulawesi, West Sumatra, West Java and Central Java.

In this research has been made a system that can identify the place of origin based on human voice using the Linear Predictive Coding (LPC) as feature extraction method and Artificial Neural Networks (ANN) Backpropagation as classification. At first the sound data is stored in a wav file, then it's analyzed with the LPC.

The LPC coefficient is obtained from the LPC analysis then it becomes the ANN input vector for training the network. The results of training in the form of changes in network weights to obtain a network that has the ability to classify. Then the network is tested by simulating it using training data and evaluation data to produce a percentage of successful recognition. The test is done with several changes in parameter values in order to obtain the highest percentage of recognition.

The result of testing shows the highest accuracy values is 80.76% using 52 training data and 52 evaluation data. The best parameters that is used are 3 hidden layers in ANN.

Keywords: *Linier Predictive Coding (LPC), Artificial Neural Network (ANN), Voice, Accent.*

1. Pendahuluan

Suara manusia memiliki ciri khas masing-masing. Pengucapan kata yang sama akan menghasilkan pola sinyal yang berbeda sesuai dengan aksen atau pelafalannya. Hal tersebut menjadi salah satu alasan adanya teknologi *speech recognition*. *Speech recognition* memungkinkan dapat mengidentifikasi asal daerah berdasarkan suara dengan membandingkan pola suara pembicara dengan pola suara kata yang sudah tersimpan dalam database.

Identifikasi suara mempunyai banyak metode yang dapat digunakan baik dalam metode ekstraksi ciri dan metode dalam membangun ciri pencocokan. Pada tugas akhir ini menggunakan metode ekstraksi ciri *Linier Predictive Coding* (LPC), hal ini dikarenakan LPC telah banyak digunakan dalam penelitian terkait identifikasi suara. Pada penelitian sebelumnya yang dilakukan oleh Ririn Kusumawati, LPC digunakan untuk menganalisis kesesuaian menulis bahasa Arab oleh penutur Indonesia dengan nilai akurasi dibawah 80% [1]. LPC pertama kali digunakan pada tahun 1978 untuk membuat alas sintesis sinyal percakapan[2]. Pada proses klasifikasi menggunakan metode Artificial Neural Network (ANN) atau Jaringan Syaraf Tiruan (JST). Pada hasil penelitian sebelumnya yang dilakukan oleh Rayani Budi Andhini, metode klasifikasi ini dilakukan untuk menganalisis dan mengimplementasi pengenalan suara menjadi teks dan menghasilkan nilai akurasi sebesar 51% [3].

Pada penelitian tugas akhir ini dilakukan pengidentifikasian asal daerah berdasarkan suara menggunakan perangkat lunak bahasa pemrograman, dengan ekstraksi ciri *Linier Predictive Coding* (LPC) dan klasifikasi Jaringan Saraf Tiruan (LPC).

2. Tinjauan Pustaka

2.1 Pengenalan Suara

Pengenalan suara merupakan proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan konversi sinyal akustik yang ditangkap oleh perangkat perekam suara. Pengenalan suara juga merupakan system yang digunakan untuk mengenali perintah kata dari suara manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer[4].

Badan Pengembangan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan (Badan Bahasa Kemendikbud) telah memetakan dan memverifikasi 652 bahasa daerah di Indonesia[5]. Dari Bahasa daerah tersebut, masing-masing memiliki dialek yang berbeda. Dialek adalah varian-varian sebuah bahasa yang sama. Varian-varian ini berbeda satu sama lain, tetapi masih banyak menunjukkan kemiripan satu sama lain sehingga belum pantas disebut bahasa-bahasa yang berbeda. Biasanya pemberian dialek adalah berdasarkan geografi, namun bisa berdasarkan faktor lain, misalkan faktor sosial. Sebuah dialek dibedakan berdasarkan kosa kata, tata bahasa, dan pengucapan.

2.2 Ekstraksi Ciri

Ekstraksi ciri suara adalah untuk mengubah gelombang suara menjadi beberapa tipe representasi parametrik yang dapat diproses. Ada banyak cara untuk merepresentasikan suara secara parametris sehingga dapat diproses lebih lanjut. Salah satunya adalah menggunakan *Linier Predictive Coding* (LPC) [6]. Metode LPC diturunkan dari ide dasar bahwa contoh sinyal ucapan pada saat ke n , $s(n)$, dapat didekati sebagai kombinasi linier dari sejumlah p contoh ucapab sebelumnya, yang dapat dinotasikan sebagai berikut:

$$s(n) \approx a_1s(n-1) + a_2s(n-2) + \dots + a_p s(n-p) \quad (2.1)$$

Sinyal ucapan oada contoh ke- n , $s(n)$ dapat didekati oleh beberapa buah contoh sebelumnya dengan bantuan koefisien prediksi a_i , koefisien a_1, a_2, \dots, a_i diasumsikan sebagai konstanta dalam blok analisis sinyal ucapan. Nilai estimasi a_i merupakan hasil terbaik jika jumlah kuadrat kesalahan antara sinyal sebenarnya sn dan sinyal perkiraan sn minimum. Ada pun tahapan yang dilakukan adalah sebagai berikut[15]:

1. *Pre-emphasis*, adalah tahap pertama dari LPC yang memfilter sinyal suara masukan. Sinyal suara frekuensi tinggi dilemahkan sehingga sinyal lebih tahan terhadap efek presisi atau pergeseran. Sistem digital yang digunakan oleh fungsi *pre-emphasis* adalah *fixed* atau *slowly adaptive*.
2. *Frame Blocking*, sinyal ucapan hasil *pre-emphasis*, $s(n)$, kemudian di blok ke dalam *frame-frame* sebanyak N sampel. *Frame-frame* yang saling berhubungan dipisahkan oleh M sampel.
3. *Windowing*, pemberian jendela (*windowing*) pada tiap *frame*. *Windowing* memiliki tujuan meminimalisasi diskontinuitas sinyal pada awal dan akhir masing-masing *frame*.

4. Analisis autokorelasi, masing-masing *frame* dari sinyal yang telah dilakukan proses *windowing*, dilakukan autokorelasi. Pada rumus 1. Nilai p adalah nilai tertinggi pada autokorelasi. Nilai p digunakan pada langkah berikutnya yaitu analisis LPC.

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n) \cdot \tilde{x}_l(n+m) \tag{2.2}$$

Dimana nilai auto korelasi yang tertinggi pada $m=p$ adalah orde dari analisa LPC.

5. Analisis LPC, merupakan proses untuk mengubah masing-masing frame dari autokorelasi $p + 1$ ke dalam bentuk parameter LPC. Parameter tersebut dapat terdiri dari koefisien LPC, koefisien refleksi, koefisien log area ratio, koefisien cepstral, atau transformasi ke bentuk set yang diinginkan.

Salah satu metode untuk melakukan hal ini ialah metode *Durbin* yang dinyatakan dalam algoritma di bawah ini [7]:

$$E^{(0)} = r(0)$$

$$k = \frac{\{r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(|i-j|)\}}{1}, 1 \leq i \leq E^{(i-1)}p \tag{2.3}$$

$$\alpha_i^{(i)} = k_i \tag{2.4}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_j^{(i-1)}, 1 \leq j \leq i - 1$$

$$E^i = (1 - k_i^2)E^{(i-1)} \tag{2.5}$$

Persamaan di atas di *rekursi* untuk $i=1,2,\dots,p$ dan penyelesaian akhirnya berupa:

$$a_m = \text{koefisien LPC} = \alpha_m^{(p)}, 1 \leq m \leq p$$

$$k_m = \text{koefisien PARCOR (koefisien pantulan)}$$

$$g_m = \text{koefisien perbandingan daerah logaritmis}$$

$$= \log \left(\frac{1 - k_m^2}{1 + k_m^2} \right)$$

6. Mengubah LPC parameter ke koefisien *cepstral*, koefisien *cepstral* dilakukan untuk memperoleh penurunan pada koefisien LPC, di mana pada parameter LPC sangat penting untuk mengubah koefisien *cepstral* agar sistem mendapatkan ciri yang tetap.

Langkah terakhir yang dilakukan pada ekstraksi ciri LPC adalah menentukan statistik *feature* agar mendapatkan ciri yang tepat pada sistem. Ekstraksi ciri merupakan metode pengambilan ciri yang didasarkan pada statistik *feature*. Pada statistik *feature* menunjukkan *speech* dan masing-masing nilai dari ekstraksi ciri. Dari nilai-nilai pada *speech* yang dihasilkan, dapat dihitung beberapa parameter ciri, antara lain adalah *mean*, *standar deviasi*, *skewness*, *variance*, *kurtosis*, dan *entropy* [7].

a. *Mean* (μ)

$$\mu = \sum_n f_n p(f_n) \tag{2.6}$$

Di mana f_n merupakan suatu nilai intensitas keabuan, sementara $p(f_n)$ menunjukkan nilai *speech* (probabilitas kemunculan intensitas tersebut pada suara).

b. Variance (σ^2)

Menunjukkan variasi elemen pada *speech* grafik dari suatu suara.

$$\sigma^2 = \sum_n (f_n - \mu)^2 p(f_n) \quad (2.7)$$

c. Standar Deviasi (s)

Standar deviasi adalah nilai yang menggambarkan seberapa besar perbedaan antar data (hasil pengukuran) atau persebaran datanya terhadap nilai rata-rata.

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.8) \text{ d.}$$

Skewness (σ^3)

Menunjukkan tingkat kemencengan relatif kurva *speech* grafik dari suatu suara.

$$\sigma^3 = \frac{1}{\sigma^3} \sum_n (f_n - \mu)^3 p(f_n) \quad (2.9)$$

e. Kurtosis (σ^4)

Menunjukkan tingkat keruncingan relatif kurva *speech* grafik dari suatu suara.

$$\sigma^4 = \frac{1}{\sigma^4} \sum_n (f_n - \mu)^4 p(f_n) - 3 \quad (2.10)$$

f. Entrophy (H)

Menunjukkan ukuran ketidakaturan bentuk dari suatu suara.

$$H = - \sum_n p(f_n) \cdot \log p(f_n) \quad (2.11)$$

2.3 Jaringan Saraf Tiruan

Jaringan Syaraf Tiruan (JST) adalah sistem adaptif yang dapat mengubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun intrernal yang mengalir melalui jaringan tersebut [8]. ANN juga digunakan untuk mengelola hubungan antara input dan output sehingga menemukan pola-pola data.

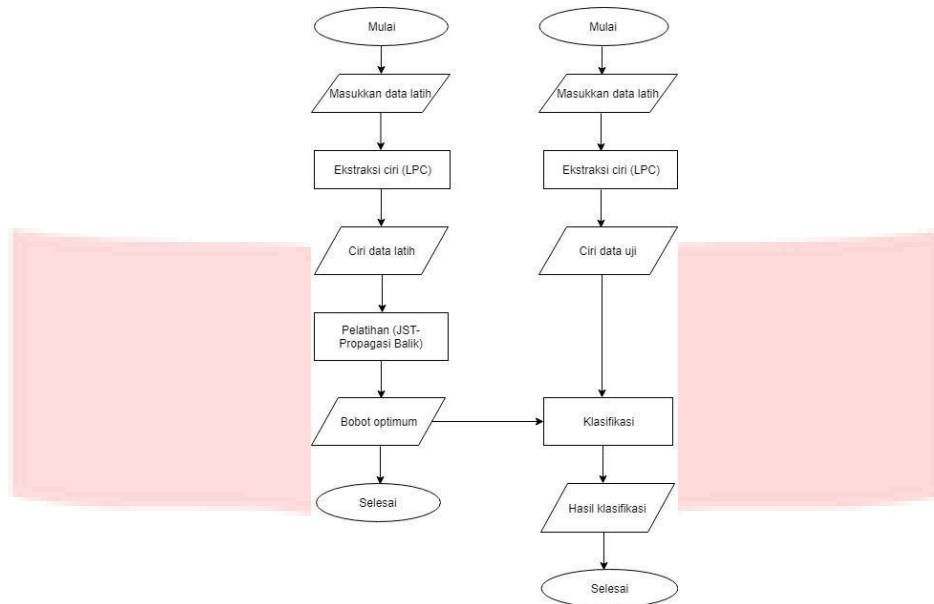
JST *backpropagation* merupakan metode *supervised learning*, di mana telah ditentukan pola *output* yang diinginkan. Arsitektur yang digunakan JST bisa berupa *single layer net* atau *multilayer net*. Proses latih dalam metode JST *Backpropagation* melibatkan 3 (tiga) tahap, yaitu [9]:

1. Tahap *feedforward*, untuk melatih pattern data input.
2. Tahap kalkulasi dan *backpropagation* dari *error*.
3. Tahap penyesuaian bobot.

3. Perancangan Sistem

3.1 Desain Sistem

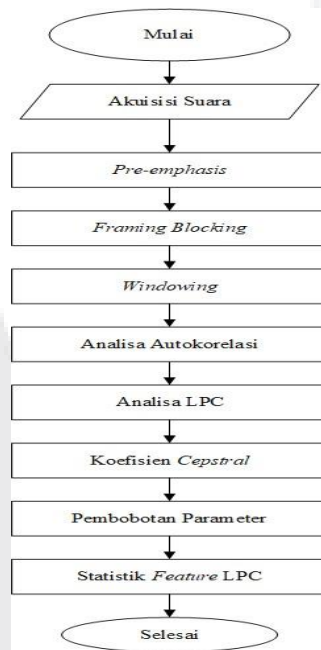
Diagram alir keseluruhan sistem pada gambar 3.1 menjelaskan bahwa suara manusia yang dijadikan data latih dilakukan proses ekstraksi ciri LPC untuk mendapatkan ciri dari data latih tersebut. Kemudian dilakukan pelatihan dengan data target yang merepresentasikan kelas dengan hasil berupa bobot. Pada data uji dilakukan ekstraksi ciri yang hasil tersebut menjadi masukan terhadap bobot untuk proses klasifikasi. Dari proses ini menghasilkan data hasil klasifikasi dan nilai akurasi.



Gambar 3.1 Diagram alir keseluruhan sistem

3.2 Ekstraksi Ciri

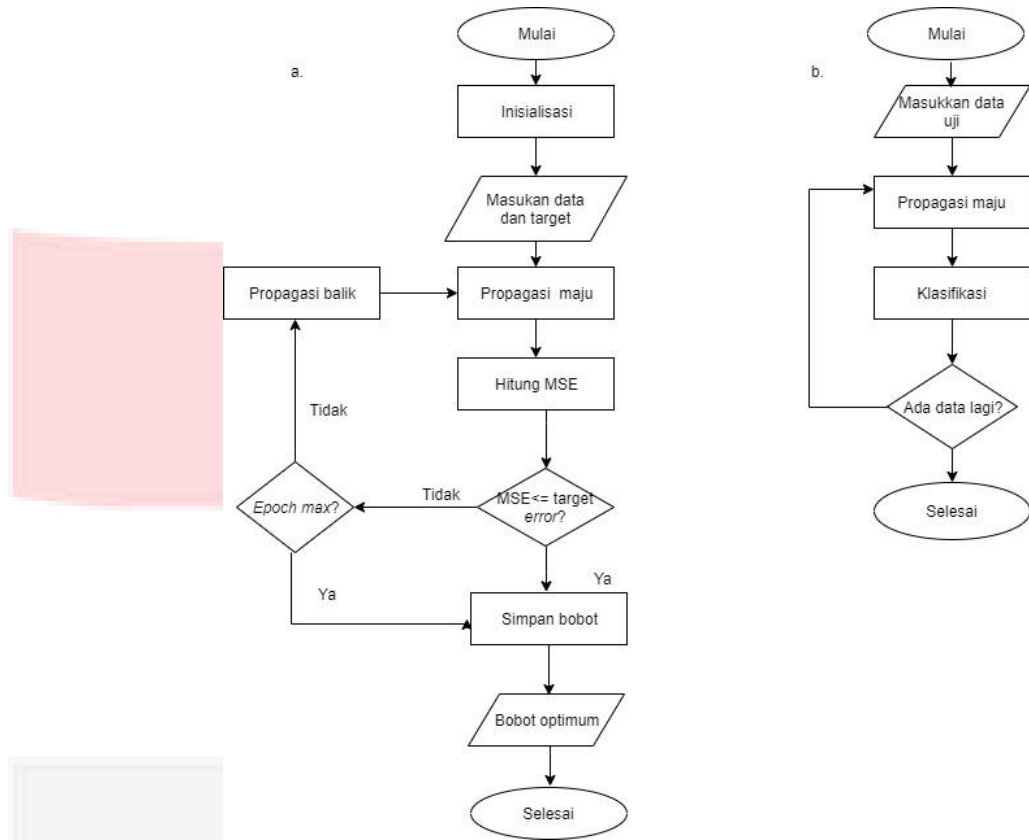
LPC adalah salah satu metode yang digunakan untuk mempresentasikan sinyal atau lebih tepatnya LPC sebagai pem-filter-an suara yang dijadikan ciri khas untuk membedakan suara yang sudah ada didalam *database* dengan suara yang di *input*. Berikut merupakan tahapan dari ekstraksi ciri menggunakan metode LPC:



Gambar 3.1 Tahapan Ekstraksi Ciri

3.3 Klasifikasi

Proses pelatihan membutuhkan target pelatihan yang direpresentasikan sebagai kelas pengujian.



Gambar 3.3 (a) Diagram alir pelatihan dan pengujian dan (b) JST Propagasi Balik

3.3.1 Pelatihan JST

Pelatihan JST dari propagasi maju yang meneruskan informasi dari *input* hingga *output* kemudian dilakukan propagasi mundur untuk memperbaiki bobot-bobot agar didapatkan bobot optimum.

3.3.2 Pengujian JST

Pengujian dilakukan terhadap data yang sebelumnya telah dilakukan pelatihan. Pengujian dilakukan untuk mengidentifikasi ciri dalam kelas-kelas pada *database*.

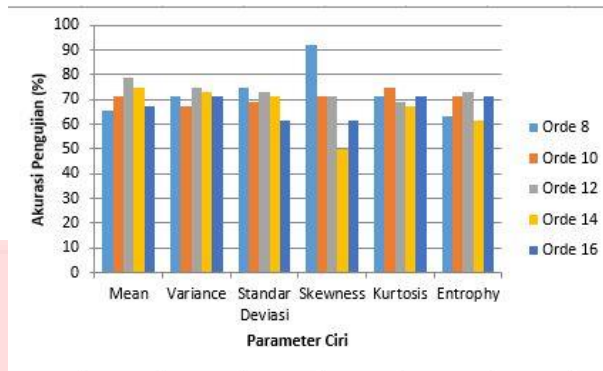
3.3.3 Parameter Keberhasilan

Pada penelitian tugas akhir ini, tingkat keberhasilan dihitung berdasarkan tingkat akurasi dengan persamaan matematis yaitu:

$$Akurasi = \frac{\sum Prediksi Benar}{\sum Prediksi Target} \times 100 \tag{3.1}$$

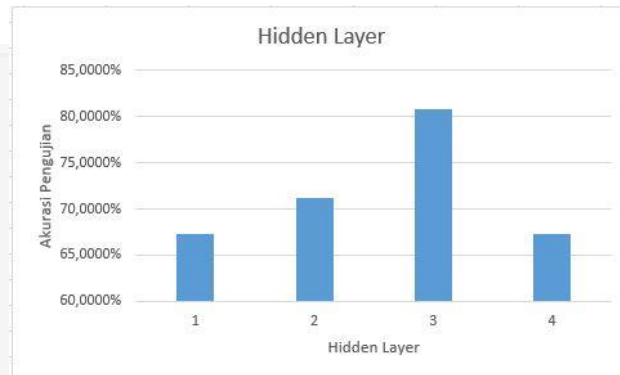
4. Analisis

Pada skenario pertama didapatkan parameter ciri paling tepat pada proses ekstraksi ciri berdasarkan nilai akurasi pengujian adalah ciri mean dengan orde 12. Pada Gambar4.1 menunjukkan kenaikan akurasi pada parameter ciri *mean*, dengan akurasi 78,8% dengan waktu komputasi 0.02 detik.



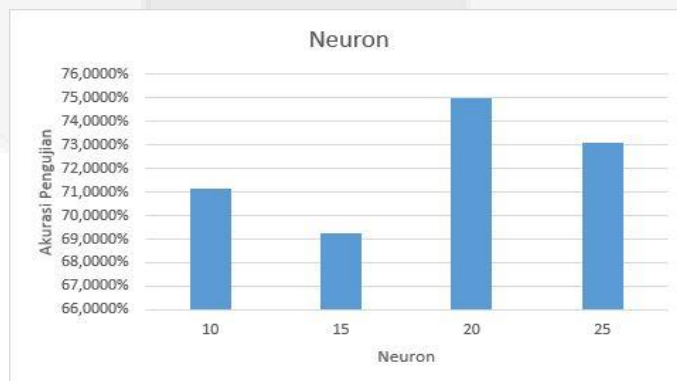
Gambar 4.1 Grafik nilai akurasi terhadap parameter ciri

Pada skenario pengujian dengan jumlah hidden layer didapatkan hasil akurasi terbaik yang terdapat pada 3 *hidden layer* dengan nilai akurasi 80.76% dan waktu komputasi 0.04 detik. Pada gambar 4.2 menunjukkan kenaikan nilai akurasi dari 2 *hidden layer* ke 3 *hidden layer*. Semakin banyak jumlah *hidden layer* maka semakin banyak node yang terhubung antar *neuron*, semakin banyak jumlah *node* maka semakin banyak bobot yang digunakan yang mengakibatkan jaringan dapat menyimpan pengetahuan yang lebih baik mengenai pola ciri dari proses pelatihan. Semakin banyak *hidden layer* maka akan memperpanjang tahap pembelajaran sehingga *network* yang dihasilkan mampu memetakan input dan target lebih tepat. Namun sistem dapat mengalami *overfitting* yaitu jaringan memproses terlalu banyak informasi melebihi kapasitas sehingga nilai akurasi menurun.



Gambar 4.2 Grafik nilai akurasi terhadap *hidden layer*

Pada skenario pengujian dengan jumlah *neuron* dengan menggunakan jumlah *hidden layer* paling tepat dengan menggunakan hasil terbaik dari skenario pengujian kedua. Pada gambar 4.3 menunjukkan grafik pengujian dengan jumlah *neuron* didapatkan nilai akurasi 75% dan waktu komputasi 0.07 detik. Semakin banyak *neuron* maka semakin baik fungsi aktivasi. Fungsi aktivasi mempengaruhi *hyperplane* dalam menyesuaikan bentuk data masukan atau memetakan kelas-kelas pada data.



Gambar 4.3 Grafik nilai akurasi terhadap *neuron*

5. Kesimpulan

Perancangan sistem yang dibuat mampu mendeteksi asal daerah dengan menggunakan metode ekstraksi ciri *Linier Predictive Coding* (LPC) dan klasifikasi Jaringan Saraf Tiruan (JST) dengan metode Propagasi Balik yang didapatkan nilai akurasi 80.76% dengan orde LPC 12 dan menggunakan parameter ciri *mean*, 3 *hidden layer* sebagai parameter di JST-PB.

Pemakaian jumlah *hidden layer* pada JST-PB mempengaruhi kinerja sistem. Dimana semakin banyak jumlah *hidden layer* maka semakin banyak *node* yang terhubung antar *neuron* dan semakin banyak jumlah *node* maka semakin banyak bobot yang digunakan yang mengakibatkan jaringan dapat menyimpan pengetahuan yang lebih baik mengenai pola ciri dari proses pelatihan. Akan tetapi sistem dapat mengalami *overfitting*. Saat menggunakan 1 *hidden layer* mendapatkan akurasi 63.30%, saat 2 *hidden layer* mendapatkan akurasi yang lebih baik bernilai 71.15% dengan waktu komputasi 0.0474 detik. Saat menggunakan 3 *hidden layer* mendapatkan nilai akurasi 80.7692% dengan waktu komputasi lebih efisien bernilai 0.0456 detik. Oleh karena itu, menambahkan jumlah *hidden layer* dapat meningkatkan nilai akurasi melalui proses trial and error.

Jumlah *neuron* yang digunakan berpengaruh terhadap kinerja sistem. Semakin banyak *neuron* maka semakin baik fungsi aktivasi. Fungsi aktivasi mempengaruhi *hyperplane* dalam menyesuaikan bentuk data masukan atau memetakan kelas datanya. Kenaikan nilai akurasi dari 15 *neuron* dengan akurasi 69.2308% ke 20 *neuron* dengan akurasi 75%. Saat penggunaan 25 *neuron* nilai akurasi menurun. Oleh karena itu, semakin banyak jumlah *neuron* yang digunakan tidak menjamin nilai akurasi .

Daftar Pustaka:

- [1] M. A. E. F. Rochman, "Rancang bangun kamus visual bahasa inggris tentang peternakan berbasis flash dengan menggunakan voice recognition," Bimasakti.
- [2] A. Suroso, Y. Fitri, S. Retnowaty, and N. Nurkhamdi, "Aplikasi pengenalan ucapan dengan ekstraksi ciri mel-frequency cepstrum coefficients (mfcc) dan jaringan syaraf tiruan (jst) propagasi balik untuk buka dan tutup pintu," Jurnal Komputer Terapan, vol. 1, no. 2, Dec. 2015. [Online]. Available: //jurnal.pcr.ac.id/index.php/jkt/article/view/63
- [3] A. Wibowo, A. Hidayanto, and A. Ajulian, "Pengenalan huruf jawa tulisan tangan menggunakan jaringan saraf tiruan perambatan balik dengan fuzzy feature extraction," Semarang: Jurusan Teknik Elektro Universitas Diponegoro, 2016.
- [4] R.Kusumawati, "Metodelinearpredictivecoding(lpc)padaklasifikasihidden markov model (hmm) untuk kata arabic pada penutur indonesia," MATICS, vol. 8, no. 1, pp. 32–35, 2016.
- [5] R. R. Juniansyah, R. Magdalena, and L. Novamizanti, "Perancangan sistem pengenalan suara dengan metode linear predictive coding," eProceedings of Engineering, vol. 4, no. 1, 2017.
- [6] A. C. Mobonguni, E. Susanto, and F. Y. Suratman, "Perancangan dan implementasi pengenalan ucapan instruksi pada sistem kendali dengan metoda klasifikasi jaringansyaraf tiruan propagasi balik," eProceedings of Engineering, vol. 2, no. 3, 2015.
- [7] K. Anam, "Pengenalan suara manusia menggunakan metode linier predictive coding (lpc)," Ph.D. dissertation, Universitas Islam Negeri Maulana Malik Ibrahim, 2013.
- [8] "Badan Bahasa Petakan 652 Bahasa Daerah di Indonesia," <https://www.kemdikbud.go.id/main/blog/2018/07/badan-bahasa-petakan-652-bahasa-daerah-di-indonesia>, [Online; accessed 19-Februari-2019].
- [9] A. Setiawan, A. Hidayatno, and R. R. Isnanto, "Aplikasi pengenalan ucapan dengan ekstraksi mel-frequency cepstrum coefficients (mfcc) melalui jaringan syaraf tiruan (jst) learning vector quantization (lvq) untuk mengoperasikan kursor komputer," Ph.D. dissertation, Diponegoro University, 2012.
- [10] D. Puspitaningrum, "Pengantar jaringan syaraf tiruan," 2006.
- [11] J. Julpan, E. B. Nababan, and M. Zarlis, "Analisis fungsi aktivasi sigmoid biner dan sigmoid bipolar dalam algoritma backpropagation pada prediksi kemampuan siswa," Jurnal Teknovasi: Jurnal Teknik dan Inovasi, vol. 2, no. 1, pp. 103–116, 2018.
- [12] F. D. Syahfitri, R. Syahputra, and K. T. Putra, "Implementation of backpropagation artificial neural network as a forecasting system of power transformer peak load at bumi ayu substation," Journal of Electrical Technology UMY, vol. 1, no. 3, pp. 118–125, 2017.
- [13] Z. F. Ramli, I. Wijayanto, and S. Hadiyoso, "Deteksi kondisi konsentrasi berdasarkan sinyal eeg dengan stimulasi menghafal al-quran," eProceedings of Engineering, vol. 5, no. 3, 2018.
- [14] S. N. Endah, S. Adhy, S. Sutikno, and R. Akbar, "Automatic speech recognition for indonesian using linear predictive coding (lpc) and hidden markov model (hmm)," 2015.