



Verifikasi Suara menggunakan Jaringan Syaraf Tiruan dan Ekstraksi Ciri Mel Frequency Cepstral Coefficient

Andi Kurniawan ^{a*}

^a Universitas Sultan Fatah Demak

Naskah Diterima : 20 Juni 2016; Diterima Publikasi : 6 Mei 2017

DOI: 10.21456/vol7iss1pp32-38

Abstract

Voice recording is an important part of the evidence for the suspect, so it is necessary to verify the voice suspects to prove the allegations of the suspect. The research aims to develop a voice verification system using artificial neural networks and extraction characteristics mel frequency cepstral coefficient. As the input data analyzed is the data of the unrecognized voice recorder of the owner and the recorded data of the sound that the owner has known as the comparison data. Data input is processed by feature extraction consisting of framing, windowing, fast Fourier transform, mel frequency wrapping, discrete cosine transform resulting in mel-frequency wrapping coefficient. The mel frequency wrapping coefficient of each frame in each input voice, is used as input on pattern recognition using artificial neural networks. The results of artificial neural networks are analyzed using decision logic to get a decision whether these two voices are the same or not. The output of the system is a decision that the tested sound is the same as or not with a voice comparison. Based on the level of compatibility of the test data produces a voice verification system with mel-frequency wrapping and artificial neural networks have a rate of 96% accuracy. The accuracy of the voice verification system can be an option to help resolve the issues in verification of voice recordings.

Keywords: Voice Verification System; Mel Frequency Cepstral Coefficient; Mel-Frequency Wrapping; Artificial Neural Network

Abstrak

Rekaman suara merupakan bagian penting sebagai barang bukti bagi tersangka, sehingga perlu dilakukan verifikasi terhadap suara tersangka untuk membuktikan dakwaan terhadap tersangka. Penelitian ini bertujuan untuk mengembangkan sistem verifikasi suara menggunakan jaringan syaraf tiruan dan ekstraksi ciri *mel frequency cepstral coefficient*. Sebagai input data yang dianalisis merupakan data rekaman suara yang tidak dikenali pemiliknya dan data rekaman suara yang sudah diketahui pemiliknya sebagai data pembanding. Input data diproses dengan ekstraksi ciri yang terdiri dari framing, windowing, fast Fourier transform, mel frequency wrapping, discrete cosine transform menghasilkan koefisien mel frequency wrapping. Koefisien mel frequency wrapping dari setiap frame pada masing-masing suara masukan, digunakan sebagai masukan pada pengenalan pola menggunakan jaring syaraf tiruan. Hasil dari jaringan syaraf tiruan dianalisis menggunakan logika keputusan untuk mendapatkan keputusan apakah kedua suara tersebut sama atau tidak. Keluaran dari sistem berupa keputusan bahwa suara yang diujikan sama atau tidak dengan suara pembanding. Berdasarkan tingkat kecocokan dari data uji menghasilkan sistem verifikasi suara dengan mel frequency wrapping dan jaring syaraf tiruan mempunyai tingkat keakuratan 96%. Akurasi sistem verifikasi suara dapat menjadi salah satu pilihan untuk membantu mengatasi permasalahan dalam verifikasi rekaman suara.

Kata kunci: Sistem Verifikasi Suara; *Mel Frequency Cepstral Coefficient*; *Mel Frequency Wrapping*; Jaringan Saraf Tiruan

1. Pendahuluan

Saat ini, penggalan informasi yang terdapat pada suara menjadi topik yang sangat menarik. Salah satunya yaitu menggunakan suara untuk mengenali pembicaraanya karena suara yang diucapkan dapat

membawa informasi mengenai identitas pembicaraanya (Pal dan Saha, 2015). Hal ini disebabkan karena setiap suara manusia memiliki frekuensi yang khas, frekuensi tersebut biasa disebut frekuensi dominan. Dalam keadaan pengucapan yang normal frekuensi dominan setiap manusia selalu sama.

Penerapan pengenalan suara untuk mendapatkan identitas pembicara dibagi dalam dua tugas yang berbeda yaitu identifikasi dan verifikasi. Perbedaan utama dalam kedua kelompok tugas tersebut terletak pada tujuannya. Tujuan identifikasi suara adalah untuk menentukan identitas pembicaranya. sedangkan verifikasi adalah menerima atau menolak identitas dari sampel suara. Penerapan identifikasi sudah pernah diteliti oleh peneliti lain yang menggunakannya untuk keperluan absensi kehadiran (Riyanto, 2013). Untuk verifikasi, banyak digunakan untuk perbankan dan forensik.

Verifikasi suara dalam bidang forensik juga sudah digunakan di Indonesia. Seperti pada beberapa kasus korupsi yang ada di Indonesia yang menggunakan bukti berupa rekaman suara, sudah berhasil dibuktikan dengan verifikasi suara. Metode yang digunakan di Indonesia yaitu membandingkan rekaman suara yang digunakan sebagai bukti dengan satu atau lebih suara yang tertuduh. Penerapan metode ini membutuhkan beberapa pakar sebagai responden untuk mengalisis tingkat kesamaan suara dalam rekaman. Akan tetapi keakuratan metode ini sangat tergantung pada kondisi pakar (Boë, 2000).

Untuk dapat melakukan verifikasi dibutuhkan suatu metode untuk pengenalan pola, salah satunya Jaringan Syaraf Tiruan. Jaringan Syaraf Tiruan memiliki keunggulan, yaitu mudah dalam aplikasi dan juga memiliki hasil yang cukup bagus sehingga banyak peneliti menggunakannya. Akurasi yang cukup tinggi juga dihasilkan oleh JST dalam pengenalan pola dari suara yang representasikan kedalam teks pada penelitian yang dilakukan oleh Gülin Dede dan Murat Hüsnü Sazlı. Dalam penelitian tersebut, keakuratan Jaringan Syaraf Tiruan antara 98% hingga 100%. Sehingga dapat disimpulkan Jaringan Syaraf Tiruan merupakan metode yang tepat untuk menangani tugas pengenalan pola khususnya untuk data suara atau dapat disebut pengenalan suara (Dede dan Sazlı, 2010).

Agar metode Jaringan Syaraf Tiruan dapat bekerja dengan baik untuk mengolah data suara, data suara perlu dikonversi kedalam data berupa angka. Agar dapat mendapatkan hasil yang sesuai harapan, perlu dilakukan pemrosesan terlebih dahulu dengan proses yang dengan ekstraksi ciri untuk mendapatkan ciri khas dari data suara yang dibutuhkan dengan menganalisis. Dari beberapa fitur ekstraksi ciri MFCC merupakan yang populer, hal ini karena MFCC memiliki skema perhitungan efisien yang tersedia untuk ketahanan pada suara yang berbeda (Sahidullah dan Saha, 2012). MFCC juga mendapatkan hasil yang baik ketika digunakan dalam pengenalan suara seperti pada penelitian dilakukan oleh (Shahamiri dan Siti Salwah, 2014a) dan juga (Siniscalchi *et al.*, 2014b).

Penelitian ini bertujuan untuk mengembangkan sistem verifikasi suara menggunakan jaringan syaraf tiruan dan ekstraksi ciri *mel frequency cepstral coefficient*. Sebagai input data yang dianalisis merupakan data rekaman suara yang tidak dikenali

pemilikinya dan data rekaman suara yang sudah diketahui pemiliknya sebagai data pembanding. Yang membedakan penelitian ini dengan penelitian sebelumnya adalah metode yang diterapkan. Pada penelitian ini metode yang digunakan adalah metode Jaringan Syaraf Tiruan (JST) untuk verifikasi suara dengan ekstraksi ciri *Mel-Frequency Cepstral Coefficients* (MFCC) yang digunakan untuk keperluan forensik.

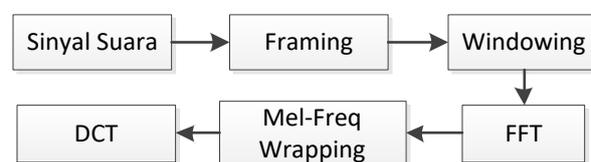
2. Kerangka teori

Teori atau yang digunakan untuk verifikasi suara dalam penelitian ini antara lain metode *mel frequency Cepstral Coefficient* (MSCC) yang digunakan untuk melakukan ekstraksi ciri pada suara dan jaringan Sjaraf tiruan (JST), Perambaan Balik yang digunakan sebagai algoritma pengenalan pola suara.

2.1. Metode *Mel frequency Cepstral Coefficient* (MFCC)

Dari semua ekstraksi ciri yang ada, MFCC merupakan yang paling populer dan telah menjadi standar dalam sistem pengenalan suara. MFCC populer juga karena skema perhitungan efisien yang tersedia untuk ketahanan pada suara yang berbeda (Sahidullah dan Saha, 2012).

Langkah kerja MFCC dalam fungsinya sebagai ekstraksi ciri dapat dilihat dalam diagram blok pada Gambar 1.



Gambar 1. Diagram blok MFCC

Framing merupakan proses pertama kali dalam perhitungan MFCC. Pada proses suara yang dianalisis dibagi menjadi beberapa frame dengan jumlah signal suara yang sama. Proses *framing* ini dilakukan terus sampai seluruh signal suara dapat diproses. Selain itu, proses ini umumnya dilakukan secara *overlapping* untuk setiap *frame*-nya. Panjang daerah *overlap* yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang *frame*. *Overlapping* dilakukan untuk menghindari hilangnya ciri atau karakteristik suara pada perbatasan perpotongan setiap *frame*.

Proses setelah *Framing* yaitu *Windowing*. Pada proses ini setiap *frame* pada proses *framing* selanjutnya dikalikan dengan jendela *Hamming* (*Windowing*) sehingga kelangsungan sisi kiri dan kanan dari *frame* dapat ditingkatkan atau dipertahankan (Ooi *et al.*, 2014). Proses ini berfungsi untuk meminimalisir diskontinuitas sinyal pada permulaan dan akhir setiap *frame*. *Hamming Window*,

$w(n)$ yang digunakan dalam MFCC didefinisikan dengan persamaan 1.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{n\pi}{N}\right), 0 \leq n \leq N - 1 \quad (1)$$

Fast Fourier Transform (FFT) digunakan untuk mengkonversi setiap frame dari domain waktu ke domain frekuensi. Hal ini perlu dilakukan karena ciri sinyal sulit untuk diamati dalam domain waktu. Fast Fourier Transform (FFT) yang didefinisikan dengan persamaan 2.

$$F_n = \sum_{k=0}^{N-1} x_k e^{-j2\pi kn}, n = 0, 1, 2, \dots, N - 1 \quad (2)$$

Dalam persepsi pendengaran manusia, sinyal suara tidak mengikuti skala linear. Jadi setiap nada dengan frekuensi dan dengan pola tertentu selanjutnya diukur dengan sebuah skala yang disebut mel. Skala mel merupakan skala frekuensi linear yang memiliki rentan dibawah 1000 Hz dan skala logaritmik di atas 1000 Hz (Ganchev *et al.*, 2005). Skala mel didefinisikan dengan persamaan 3.

$$\text{Mel}(f) = 2595 \log_{10} \left(1 + \left(\frac{f}{700} \right) \right)$$

dimana f adalah nilai dari frekuensi yang didapatkan dari perhitungan FFT.

Koefisien MFCC baru dapat dilihat dalam bentuk vektor setelah dilakukan perhitungan Discrete Cosine Transform (DCT) (Ahmed *et al.*, 1974). DCT merupakan langkah terakhir dari proses utama ekstraksi ciri MFCC. Konsep dasar dari DCT adalah mendekorelasikan *mel spectrum* sehingga menghasilkan representasi yang baik dari *property spectral local*. Pada dasarnya konsep dari DCT sama dengan alihragam fourier balik. Namun hasil dari DCT mendekati PCA (Principle Component Analysis). PCA adalah metode statistik klasik yang digunakan secara luas dalam analisis data dan kompresi. Hal inilah yang menyebabkan seringkali DCT menggantikan alihragam fourier balik dalam proses ekstraksi ciri MFCC. Persamaan yang digunakan untuk menghitung DCT yaitu Persamaan (4).

$$c(n) = \sum_{k=1}^K \log s_k \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], n = 1, 2, 3, \dots, N \quad (4)$$

2.2. Jaringan Syaraf Tiruan (JST)

Jaringan Syaraf Tiruan merupakan sistem adaptif yang dapat merubah strukturnya berdasarkan informasi yang mengalir melalui jaringan tersebut. Istilah buatan disini digunakan karena jaringan syaraf ini diimplementasikan dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran (Kusumadewi, 2003).

Secara umum JST terbagi menjadi 2 bagian yaitu pelatihan dan pengujian. Dalam pelatihan sebuah

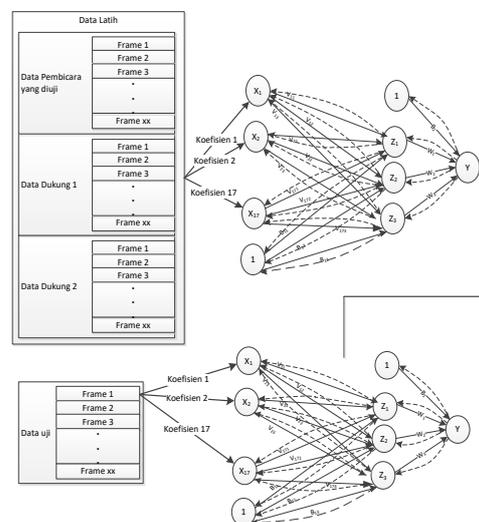
konfigurasi jaringan dapat dilatih untuk mempelajari data historis yang ada. Dengan pelatihan yang dilakukan, pengetahuan yang terdapat pada data dapat diserap dan direpresentasikan oleh nilai-nilai bobot koneksinya. Sedangkan pengujian merupakan proses pengujian ketelitian dari model yang sudah diperoleh dari model proses pelatihan.

2.3. Jaringan Syaraf Tiruan Perambatan Balik

Jaringan Syaraf Tiruan perambatan balik atau Backpropagation (BP) adalah algoritma paling populer untuk pelatihan *feed-forward artificial neural network* (FFANN), dimana bobot yang disesuaikan dengan metodologi *gradient descent* dan turunan dari fungsi kesalahan sehubungan dengan bobot yang efisien dihitung dengan prosedur yang disebut perambatan balik yang terdiri dari dua fase: transmisi maju dan transmisi mundur. Pada fase penalaran ke depan, pola pelatihan ditambahkan ke lapisan masukan, dan kemudian arus informasi merambat dari lapisan ke lapisan sampai pola keluaran diperoleh sebagai tanggapan (*response*) dari masukan (Zhang *et al.*, 2015).

2.3.1. Arsitektur JST Perambatan Balik

JST Perambatan Balik termasuk dalam JST yang memiliki banyak lapisan dengan 3 lapisan, yaitu lapisan masukan, lapisan tersembunyi dan lapisan keluaran. Arsitektur JST Perambatan Balik ditunjukkan pada Gambar 2. Dalam Gambar 2. ditunjukkan masukan dalam proses pelatihan menggunakan 3 data suara yaitu data pembicara yang diuji, data dukung 1, dan data dukung 2. Setiap data suara yang sudah melewati proses ekstraksi ciri sebelumnya menghasilkan beberapa frame dimana setiap frame terdiri dari 17 koefisien MFCC. Koefisien MFCC yang selanjutnya digunakan sebagai masukan JST, dalam arsitektur jaringan yang ditunjukkan pada Gambar 2 diwakili oleh X. Sedangkan Z adalah layer tersembunyi, dan Y adalah keluaran.



Gambar 2. Arsitektur JST perambatan balik

Dalam Gambar 2. ditunjukkan masukan dalam proses pelatihan menggunakan 3 data suara yaitu data pembicara yang diuji, data dukung 1, dan data dukung 2. Setiap data suara yang sudah melewati proses ekstraksi ciri sebelumnya menghasilkan beberapa frame dimana setiap frame terdiri dari 17 koefisien MFCC. Koefisien MFCC yang selanjutnya digunakan sebagai masukan JST, dalam arsitektur jaringan yang ditunjukkan pada Gambar 2 diwakili oleh X. Sedangkan Z adalah layer tersembunyi, dan Y adalah keluaran.

Dalam JST terdapat bobot yang selanjutnya dikalikan dengan masukan kemudian digunakan fungsi aktivasi sehingga menghasilkan keluaran dari satu neuron. Bobot antar lapisan tersembunyi dan masukan diwakili oleh V. Satu neuron masukan terhubung ke seluruh neuron layer tersembunyi. Variabel i digunakan untuk mewakili jumlah koefisien MFCC. i terdiri dari 1 hingga 17. j digunakan untuk mewakili jumlah layer tersembunyi. Variabel j terdiri 1 hingga jumlah layer tersembunyi dalam jaringan. W adalah bobot antara lapisan tersembunyi dan keluaran. B_{1j} Adalah bobot dari bias masukan ke lapisan tersembunyi. Sedangkan B_{2j} adalah bobot dari bias lapisan tersembunyi ke keluaran.

3. Metodologi

3.1. Tahapan Identifikasi

Tahapan ini meliputi identifikasi masalah, tujuan penelitian serta melakukan observasi dengan objek penelitian suara manusia sebagai sampel untuk mendapatkan hasil permasalahan yang nyata dan data-data pendukung penelitian. Pengambilan data suara yang akan dianalisis, diambil dalam situasi dimana suara pembicara paling natural. Dalam pengambilan data, suara satu orang direkam beberapa kali yaitu sebagai data latih, data uji dengan pembicaraan yang sama, data uji dengan pembicaraan yang berbeda, data uji yang diambil dari rekaman telepon.

3.2. Tahapan Perancangan Sistem

Dalam tahapan perancangan sistem ini dilakukan dalam kegiatan berupa penerapan Jaringan Syaraf Tiruan sebagai pengenalan pola dan MFCC (*Mel Frequency Cepstral Coefficient*) sebagai ekstraksi ciri untuk analisis suara pada pemrograman sistem verifikasi biometrik suara menggunakan Matlab. Adapun Kerangka Sistem tersebut ditunjukkan pada Gambar 3.

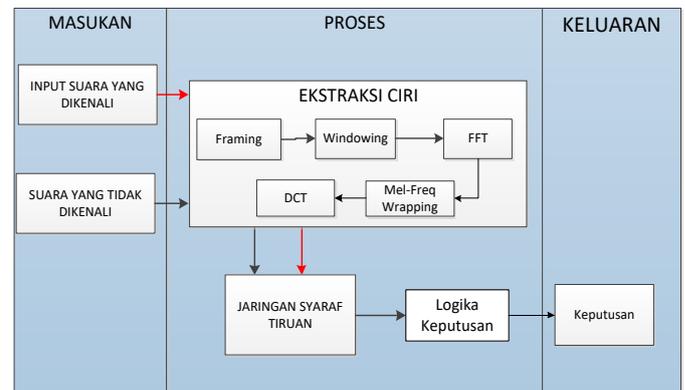
3.3. Tahapan Implementasi

Pada tahapan ini, dilaksanakannya penerapan sistem yang sudah jadi dan kemudian melakukan testing terhadap sistem dengan menggunakan data suara yang sudah dikumpulkan sebelumnya. Setelah itu, dianalisis bahwa dari hasil simulasi yang dilakukan sudah sesuai dengan tujuan penelitian apa tidak. Jika sudah sesuai akan dilanjutkan ke tahapan kesimpulan.

3.4. Tahapan Kesimpulan

Tahapan ini mengambil kesimpulan yang didapat dari tahapan implementasi.

Kerangka Sistem Informasi Verifikasi Suara menggunakan Jaringan Syaraf Tiruan dan Ekstraksi Ciri MFCC (*Mel Frequency Cepstral Coefficient*) ditunjukkan dengan Gambar 3.



Gambar 3. Kerangka Sistem

Penjelasan kerangka Sistem Informasi Verifikasi Suara menggunakan Jaringan Syaraf Tiruan dan Ekstraksi Ciri MFCC (*Mel Frequency Cepstral Coefficient*) pada Gambar 3. adalah sebagai berikut.

3.5. Masukan

Pada tahap ini pengguna sistem memasukkan data rekaman suara yang tidak dikenali pemiliknya dan data rekaman suara yang sudah diketahui pemiliknya sebagai data pembandingan.

3.6. Proses

Pada proses ini, data dari proses masukan diolah dengan ekstraksi ciri yang terdiri dari *framing*, *windowing*, *FFT*, *Mel-frequency wrapping*, *DCT* yang kemudian menghasilkan koefisien MFCC. Kemudian koefisien MFCC dari setiap frame pada masing-masing suara masukan, digunakan sebagai masukan pada pengenalan pola menggunakan Jaringan Syaraf Tiruan (JST). Dimana koefisien MFCC dari suara yang suara pembandingan sebagai data training dan suara yang tidak dikenal sebagai data uji. Selanjutnya hasil dari Jaringan Syaraf Tiruan dianalisis menggunakan logika keputusan untuk mendapatkan keputusan apakah kedua suara tersebut sama atau tidak.

3.7. Keluaran

Keluaran dari sistem tersebut yaitu keputusan bahwa suara yang diujikan sama atau tidak dengan suara pembandingan.

4. Hasil dan Pembahasan

Hasil penelitian ini adalah sistem verifikasi suara. Verifikasi suara ini membutuhkan masukan berupa suara digital yang mempunyai format *.WAV.

Tahapan akuisisi data ini mengubah suara analog menjadi digital menggunakan alat mikropon yang dihubungkan dengan laptop. Hasil dari proses ini yaitu data suara yang diambil dari 15 individu yang berbeda dan setiap individu diambil 10 kali pengambilan. Data suara yang telah diambil tersebut selanjutnya diproses dengan ekstraksi ciri untuk mendapatkan ciri khas dari setiap suara yang selanjutnya disimpan kedalam basisdata.

Metode yang digunakan sebagai ekstraksi ciri untuk suara penelitian ini yaitu menggunakan metode MFCC, MFCC merupakan salah satu tehnik ekstraksi ciri berbasis transformasi *Fourier* yang telah banyak dipakai dalam pemrosesan suara. Proses MFCC diawali dengan proses *Framing*. Pada proses ini data suara dipotong sepanjang 25 ms dan pergeseran setiap frame 10 ms (Siniscalchi, Svendsen, dan Lee, 2014) dengan *frequency sampling* 11040 Hz. Sehingga dalam satu frame terdapat 276 sampel.

Setelah melalui proses *framing*, data suara diproses dengan *windowing*. Pada proses ini nilai sampel data di setiap *frame* dikalikan dengan nilai jendela Hamming. Proses ini berfungsi untuk meminimalkan diskontinuitas sinyal setelah melewati proses *Framing*. Setelah melalui proses *windowing*, sampel data diproses dengan FFT. *Fast Fourier Transform* (FFT) digunakan untuk mengkonversi setiap frame dari domain waktu ke domain frekuensi.

Setelah melalui proses FFT, sampel data diproses dengan *Mel-Frequency Wrapping*. Pada proses ini setiap data dengan frekuensi dan dengan pola tertentu selanjutnya diukur dengan sebuah skala yang disebut *Mel*. Pada penelitian ini panjang filter yang digunakan dalam *Mel-Frequency Wrapping* sebanyak 20 filter. Hasil dari *Mel-Frequency Wrapping* ini dalam setiap *frame* terdiri dari 20 data.

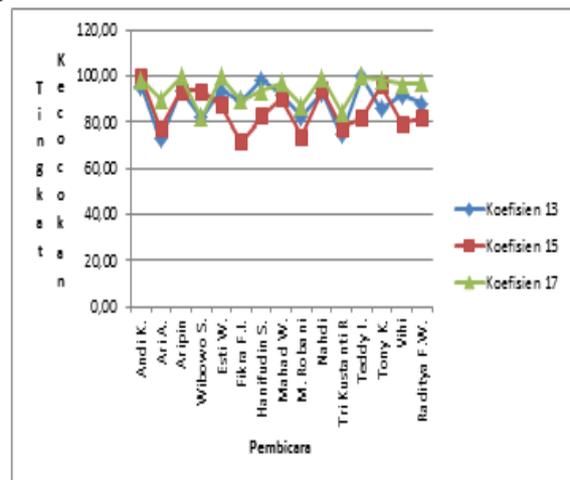
Setelah melalui proses *Mel-Frequency Wrapping*, sampel data diproses dengan DCT. Proses DCT merupakan langkah terakhir dari proses utama MFCC. Hasil dari proses ini yaitu koefisien MFCC yang dalam penelitian ini menggunakan tiga koefisien yaitu koefisien 13, 15 dan 17.

Untuk pengenalan dalam program ini menggunakan JST perambatan balik. Dalam JST tidak ada aturan baku yang dapat menetapkan arsitektur jaringan terbaik. Untuk mengetahui arsitektur jaringan terbaik, dapat diketahui dengan melakukan beberapa pengujian sampai ditemukan arsitektur dengan hasil yang optimal.

Sebelum melakukan verifikasi, Jaringan Syaraf Tiruan harus dilakukan pelatihan. Pelatihan tersebut menggunakan data yang sudah tersimpan di dalam basis data. Proses pelatihan Jaringan Syaraf Tiruan untuk verifikasi suara tidak menggunakan seluruh data suara pada basis data. Data yang digunakan dalam pelatihan yaitu data pembicara yang dipilih dan 2 data pendukung.

Setelah melakukan pelatihan, proses pengenalan pola dengan jaringan syaraf tiruan dilanjutkan dengan

pengujian. Data suara yang diekstraksi pada proses sebelumnya diuji dengan pengenalan pola Jaringan Syaraf Tiruan. Selanjutnya hasil dari pengujian dihitung tingkat kemiripannya dengan menghitung jaraknya. Jarak yang terdekat dengan pembicara selanjutnya dihitung sebagai probabilitas hipotesa pembicara sama. Sedangkan jika jarak terdekat tidak pada pembicara dihitung sebagai probabilitas hipotesa pembicara berbeda. Selanjutnya dari probabilitas hipotesa sama tersebut diukur tingkat kecocokannya. Adapun hasil keluaran Jaringan Syaraf Tiruan berupa tingkat kecocokan ditunjukkan dalam bentuk grafik pada Gambar 4.



Gambar 4. Grafik tingkat kecocokan dengan koefisien 13,15,17

Pada Gambar 4. ditunjukkan adanya kecenderungan (*trend*) menurunnya tingkat kecocokan setelah koefisien MFCC 13 ditingkat menjadi koefisien MFCC 15 dan tingkat kecocokan meningkat kembali setelah koefisien MFCCnya ditingkatkan kembali menjadi 17 koefisien. Koefisien 17 juga merupakan hasil dengan tingkat kecocokan tertinggi pada hampir seluruh pengujian.

Pengambilan keputusan pada verifikasi suara ini menggunakan algoritma rasio kemungkinan/*likelihood ratio* (LR). Dalam algoritma ini terdapat batas ambang (*threshold*) yang digunakan untuk pengambilan keputusan. Nilai batas ambang yang menunjukkan pembicara sebenarnya yaitu berkisar antar 0,2176 dan 5,2609 (Pal & Saha, 2015). Hasil dari pengambilan keputusan ini yaitu Suara cocok dan Suara tidak cocok.

Selanjutnya, pengujian verifikasi suara diukur dari tingkat akurasi dengan jumlah koefisien MFCC sebanyak 13, 15 dan 17. Pengujian ini menggunakan koefisien MFCC dengan variasi dari lapisan tersembunyi pada aplikasi yang diujikan pada data dari 15 individu yang berbeda dan setiap individu dilakukan pengujian sebanyak 10 kali. Adapun hasil pengujian verifikasi suara dengan jumlah koefisien MFCC 13, 15, dan 17 ditunjukkan oleh Tabel 1, Tabel 2 dan Tabel 3 yang selanjutnya disajikan dalam bentuk grafik yang ditunjukkan pada Gambar 5.

Tabel 1. Pengujian sistem dengan 13 Koefisien MFCC

Max Epoch	Layer Tersembunyi	Koefisien MFCC	Waktu	Jumlah Pengenalan Benar	Akurasi (%)
3000	5	13	6	128	85,33
3000	10	13	7	130	86,67
3000	15	13	6	131	87,33
3000	20	13	7	131	87,33
3000	25	13	7	135	90
3000	30	13	8	142	94,67

Berdasarkan Tabel 1. pada pengujian dengan 13 koefisien MFCC kondisi terbaik adalah lapisan tersembunyi sebanyak 15 *neuron* dengan persentase akurasi mencapai 86,67%, lapisan tersembunyi sebanyak 25 *neuron* dengan persentase akurasi mencapai 90% dan lapisan tersembunyi sebanyak 30 *neuron* dengan persentase akurasi mencapai 94,67%. Dari ketiga hasil tersebut, jaringan dengan lapisan tersembunyi sebanyak 15 *neuron* yang memiliki waktu iterasi tercepat yaitu 6 detik dengan presentase akurasi mencapai 87,3%. Sedangkan yang memiliki presentase akurasi dicapai dengan jumlah lapisan tersembunyi sebanyak 30 *neuron*, akan tetapi memiliki waktu iterasi yang lebih lama yaitu 8 detik. Jika melihat selisih waktu dan presentasi akurasi yang dicapai maka dapat dikatakan jaringan dengan jumlah lapisan tersembunyi sebanyak 30 *neuron* merupakan arsitektur jaringan terbaik untuk verifikasi menggunakan 13 koefisien MFCC.

Tabel 2. Pengujian sistem dengan 15 Koefisien MFCC

Max Epoch	Layer Tersembunyi	Koefisien MFCC	Waktu	Jumlah Pengenalan Benar	Akurasi (%)
3000	5	15	5	122	81,33
3000	10	15	5	126	84
3000	15	15	6	128	85,33
3000	20	15	7	132	88
3000	25	15	7	137	91
3000	30	15	8	142	94,67

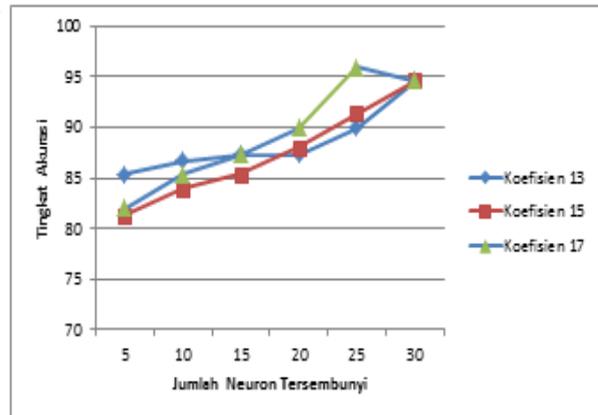
Berdasarkan Tabel 2. pada pengujian dengan 15 koefisien MFCC kondisi terbaik adalah lapisan tersembunyi sebanyak 30 *neuron* dengan persentase akurasi mencapai 94,67%, dengan waktu iterasi yang mencapai 8 detik. Jika melihat selisih waktu dan presentase akurasi yang dicapai maka dapat dikatakan jaringan dengan jumlah lapisan tersembunyi sebanyak 30 *neuron* merupakan arsitektur jaringan terbaik untuk verifikasi menggunakan 15 koefisien MFCC.

Tabel 3. Pengujian sistem dengan 17 Koefisien MFCC

Max Epoch	Layer Tersembunyi	Koefisien MFCC	Waktu	Jumlah Pengenalan Benar	Akurasi (%)
3000	5	17	5	122	82
3000	10	17	6	126	85,33
3000	15	17	6	128	87,33
3000	20	17	7	132	90
3000	25	17	7	144	96
3000	30	17	8	142	94,67

Berdasarkan Tabel 3. pada pengujian dengan 17 koefisien MFCC kondisi terbaik adalah lapisan

tersembunyi sebanyak 25 *neuron* dengan persentase akurasi mencapai 96%, dengan waktu iterasi yang mencapai 7 detik. Jika melihat selisih waktu dan presentase akurasi yang dicapai maka dapat dikatakan jaringan dengan jumlah lapisan tersembunyi sebanyak 25 *neuron* merupakan arsitektur jaringan terbaik untuk verifikasi menggunakan 17 koefisien MFCC.



Gambar 5. Grafik pengujian verifikasi suara

Pada Gambar 5. ditunjukkan adanya kecenderungan (*trend*) meningkatnya tingkat akurasi diikuti dengan ditingkatkannya jumlah parameter neuron tersembunyi. Pada pengujian menggunakan 13 koefisien MFCC dan 15 koefisien MFCC, akurasi terbaik diperoleh dari jumlah neuron tersembunyi terbesar yaitu 30. Sementara itu untuk koefisien MFCC 17, hasil terbaik yaitu dengan persentase 96% diperoleh dari jumlah neuron tersembunyi 25. Justru jika jumlah tersembunyi ditambah menjadi 30 akurasi menurun menjadi 94,67%.

5. Kesimpulan

Dari penelitian menggunakan Jaringan Syaraf Tiruan dan ekstraksi ciri MFCC (*Mel Frequency Cepstral Coefficient*) untuk verifikasi suara diperoleh kesimpulan bahwa pada pengujian JST perambatan balik terlihat adanya kecenderungan (*trend*) meningkatnya tingkat akurasi diikuti dengan ditingkatkannya jumlah parameter *neuron* tersembunyi. Pada pengujian menggunakan 13 koefisien MFCC dan 15 koefisien MFCC, akurasi terbaik diperoleh dari jumlah *neuron* tersembunyi terbesar yaitu 30. Sementara itu untuk koefisien MFCC 17, hasil terbaik yaitu dengan persentase 96% diperoleh dari jumlah *neuron* tersembunyi 25. Justru jika jumlah tersembunyi ditambah menjadi 30 akurasi menurun menjadi 94,67%. Hasil pengujian dengan program, tidak semua hasil verifikasi suara dapat mencapai tingkat kecocokan diatas 90% sehingga diperlukan penelitian lebih lanjut dengan kombinasi metode yang lain untuk mencapai tingkat kecocokan diatas 90% di seluruh pengujian.

Daftar Pustaka

- Ahmed, N., Natarajan, T., Rao, K. R., 1974. Discrete Cosine Transform. *Computers, IEEE Transactions on*, C-23(1), 90–93.
- Boë, L. J., 2000. Forensic voice identification in France. *Speech Communication*, 31(2), 205–224.
- Dede, G., Sazlı, M. H., 2010. Speech recognition with artificial neural networks. *Digital Signal Processing*, 20(3), 763–768.
- Ganchev, T., Ganchev, T., Fakotakis, N., Fakotakis, N., Kokkinakis, G., & Kokkinakis, G., 2005. Comparative evaluation of various MFCC implementations on the speaker verification task. *In Proc. of the SPECOM-2005*, 191–194.
- Kusumadewi, S., 2003. *Artificial Intelligence*, Penerbit Graha Ilmu, Yogyakarta.
- Ooi, C. S., Seng, K. P., Ang, L. M., Chew, L. W., 2014. A new approach of audio emotion recognition. *Expert Systems with Applications*, 41(13), 5858–5869.
- Pal, M., Saha, G., 2015. On robustness of speech based biometric systems against voice conversion attack. *Applied Soft Computing*, 30, 214–228.
- Riyanto, E., 2013. Speaker Recognition System with MFCC Feature Extraction and Neural Network Backpropagation. *ICISBC*, 62–66.
- Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication*, 54(4), 543–565.
- Shahamiri, S. R., Binti Salim, S. S., 2014a. Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach. *Advanced Engineering Informatics*, 28(1), 102–110.
- Shahamiri, S. R., Binti Salim, S. S., 2014b. Real-time frequency-based noise-robust Automatic Speech Recognition using Multi-Nets Artificial Neural Networks: A multi-views multi-learners approach. *Neurocomputing*, 129, 199–207.
- Siniscalchi, S. M., Svendsen, T., Lee, C.-H., 2014. An artificial neural network approach to automatic speech processing. *Neurocomputing*, 140, 326–338.
- Zhang, J., Ji, N., Liu, J., Pan, J., Meng, D., 2015. Enhancing performance of the backpropagation algorithm via sparse response regularization. *Neurocomputing*, 153, 20–40.