

STEGANALISIS SINYAL WICARA BERFORMAT .WAV MENGUNAKAN KOMBINASI METODE *MEL-FREQUENCY CEPSTRAL COEFFICIENT* (MFCC) DAN LINEAR DISCRIMINANT ANALYSIS (LDA)

Alifia Putri Anantha¹, Bambang Hidayat², Nur Andini³

^{1, 2, 3}Fakultas Teknik Elektro, Universitas Telkom, Bandung 40257, Indonesia

¹ananthafia@student.telkomuniversity.ac.id, ²bhidayat@telkomuniversity.ac.id,

³nurandini@telkomuniversity.ac.id

Abstrak

Dalam penelitian ini, telah dirancang sebuah sistem yang dapat melakukan steganalisis suatu berkas sinyal wicara berformat .wav, dengan menganalisis karakteristik ciri suatu *file speech*. Metode yang digunakan untuk mengekstraksi ciri sinyal wicara adalah *Mel-Frequency Cepstral Coefficient* (MFCC), sedangkan untuk menyeleksi ciri digunakan metode *Linear Discriminant Analysis* (LDA) metode yang digunakan untuk menyeleksi ciri adalah *Linear Discriminant Analysis* (LDA). Hasil keluaran dari sistem ini berupa kondisi yang menyatakan berkas sinyal wicara merupakan berkas sinyal wicara asli, tersisip pesan tersembunyi, atau tersisip *noise*. *Support Vector Machine* (SVM) digunakan untuk pengklasifikasian dan performansi sistem terbaik yang didapatkan adalah 82,86%. Kondisi tersebut diraih oleh penggunaan *hamming window*, *overlapping*, panjang *frame* 30 ms, dan *mel-filter bank* yang dibangun berjumlah 20 *mel-filter bank*.

Kata Kunci: steganalisis, *speech signal*, *wav*, *mel-frequency cepstral coefficient*, *linear discriminant analysis*

Abstract

In this research, it has been designed a system that can steganalyze a .wav format speech file, by analyzing the values or characteristics of the speech signal. In order to know characteristics of the signal, *Mel-Frequency Cepstral Coefficient* (MFCC) and *Linear Discriminant Analysis* (LDA) were used. The system output is capable to differentiate whether the speech signals is original, contains hidden message, or contains noise. *Support Vector Machine* (SVM) was used for classification process and the capability of classification of this system is 82.86%. This condition is achieved by the use of *hamming window*, *overlapping*, 30 ms frame length, and 20 bank mel-filters.

Key Words: steganalysis, *speech signal*, *wav*, *mel-frequency cepstral coefficient*, *linear discriminant analysis*

1. Pendahuluan

Di dunia yang kini serba digital, pertukaran informasi/data sangatlah mudah dan cepat. Dibalik berkembangnya kemudahan pertukaran data, muncul pula beragam teknik yang mampu menyisipkan atau menyembunyikan pesan rahasia ke dalam suatu media, sehingga keberadaan pesan tersebut tidak dapat diketahui oleh banyak pihak. Metode ini disebut steganografi [1]. Permasalahan yang dihadapi adalah, tidak selamanya pesan rahasia yang disisipkan bersifat baik. Maka dari itu diperlukan, suatu upaya untuk mengawasi pendistribusian pesan yang dapat mendeteksi ada tidaknya pesan rahasia yang disembunyikan ke dalam suatu media. Teknik tersebut dinamakan steganalisis [2].

Salah satu media komunikasi digital yang populer di kalangan masyarakat adalah audio berupa sinyal

wicara seperti *voice note*, sehingga kemungkinan adanya pendistribusian pesan rahasia melalui media audio sinyal wicara ini sangatlah besar. Selain itu, perbedaan antara sinyal wicara yang asli dengan yang sudah tersisipi pesan rahasia pun tidak mudah diketahui.

Penelitian ini membahas perancangan sistem steganalisis pada berkas sinyal wicara (*speech file*) berformat .wav menggunakan metode *Mel-Frequency Cepstral Coefficient* (MFCC) dan *Linear Discriminant Analysis* (LDA). MFCC dipilih karena metode ini merupakan metode yang paling populer digunakan untuk mengekstraksi *speech file* [3][3] dan LDA dipilih karena memiliki *error* yang kecil [4]. *Support Vector Machine* (SVM) digunakan dalam proses klasifikasi, karena dalam penentuan jaraknya menggunakan *support vector* sehingga proses komputasi menjadi cepat [5].

MFCC sebelumnya pernah digunakan pada penelitian [1] dan LDA sendiri digunakan pada penelitian [6]. Keduanya sama-sama merancang sistem steganalisis audio dengan akurasi masing-masing 50% [8] dan 88,33% [16]. Penelitian ini akan menggunakan kombinasi dari kedua metode tersebut untuk mengetahui distribusi nilai dari ciri berkas sinyal wicara yang dideteksi. Dengan mengetahui nilai-nilai tersebut, maka dapat diketahui perbedaan berkas sinyal wicara asli dengan yang sudah disisipi pesan rahasia, atau bahkan hanya terkena *noise*.

2. Dasar Teori

2.1 Steganografi

Secara umum, steganografi biasa didefinisikan sebagai seni menyembunyikan informasi rahasia dalam pesan lain yang tampaknya tidak berbahaya, sehingga mencegah adanya tindakan pendeteksian pesan tersembunyi [7]. Dibandingkan teknik enkripsi lain seperti kriptografi, steganografi merupakan teknik yang lebih efektif untuk meningkatkan keamanan pesan selama proses *transfer data* [8]. Menggunakan teknik kriptografi, kecurigaan pihak ketiga terhadap pesan yang dikirimkan akan tetap timbul, namun tidak jika menggunakan teknik steganografi [2]. Ada dua komponen utama yang dibutuhkan steganografi, yaitu media *carrier* dan data rahasia yang ingin disembunyikan. Media *carrier* berperan sebagai wadah penampung untuk menyisipkan pesan rahasia. Media *carrier* yang digunakan biasanya berupa media penyimpanan seperti teks, gambar, audio, dan video. Proses steganografi dapat dilihat pada Gambar 1.

2.2 Steganalisis

Pengertian steganalisis mengacu pada seni dan ilmu pengetahuan yang membedakan antara *stego object* dengan *cover object* atau bisa disebut *stego file* dengan *non-stego file* [2]. Maka dari itu, steganalisis dapat dimanfaatkan untuk mengetahui ada atau tidaknya pesan tersembunyi yang terkandung di dalam suatu media, dengan melihat adanya perubahan-perubahan ciri yang terjadi pada media tersebut.

Steganalisis dapat dilakukan pada berbagai jenis media seperti teks, audio/suara, gambar, dan video [2]. Sehebat apapun metode yang digunakan, jejak statistika penyisipan pasti akan tetap terekam. Sehingga, steganalisis pada umumnya dilakukan dengan cara membandingkan media yang menjadi *suspect* dengan media yang asli. Apabila ditemukan perbedaan diantara keduanya maka dapat ditarik kesimpulan bahwa mungkin terdapat pesan rahasia di dalamnya. Tahapan metode global dalam steganalisis adalah sebagai berikut [6][9]:

1. Membersihkan *noise* pada media (*de-noising*).
2. Mengekstraksi *feature-feature* yang ada.

3. Menganalisa perbedaan atau perubahan yang terjadi antara *stego file* dengan media aslinya (*non-stego file*) yang dapat dilihat dari karakteristik cirinya.
4. Menganalisa ciri yang telah diperoleh.
5. Mengklasifikasikan *stego file* dan *file* asli dengan melihat pola yang digunakan.

2.3 Noise

Pengertian *noise* menurut Kamus Besar Bahasa Inggris-Amerika adalah suara atau bunyi yang keras, tidak menyenangkan, tidak terduga, atau tidak diinginkan [13]. Menurut ilmu fisika, *noise* didefinisikan sebagai sebuah gangguan acak, terus menerus, dan mengaburkan kejelasan sinyal [10]. *Noise* dikarakterisasikan melalui waktu dan frekuensi. Domain waktu, *noise* dapat berupa *continuous* dan *discontinuous*. Sementara dalam frekuensi, *noise* dapat dikategorikan berupa *broad band* dan *narrow band* [11].

Bergantung pada karakteristik frekuensi, spektrum atau waktunya, proses *noise* dikelompokkan menjadi beberapa kategori, yaitu *band-limited white noise*, *white noise*, *colored noise*, *narrowband noise*, *transient noise pulses*, dan *impulsive noise* [12].

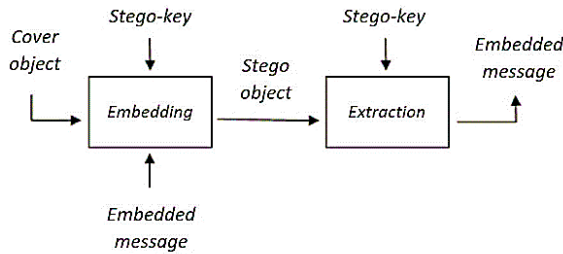
2.4 Mel-Frequency Cepstral Coefficient (MFCC)

Speech signal atau sinyal wicara terdiri dari nada (*tone*) dengan frekuensi yang berbeda-beda dan dapat didefinisikan secara subjektif pada skala non linier, skala ini dikenal sebagai skala Mel. Skala Mel adalah skala logaritmik frekuensi yang didasarkan pada persepsi *pitch* atau suara manusia [13]. *Mel-Frequency Cepstral Coefficients* (MFCC) adalah salah satu teknik ekstraksi fitur paling populer yang digunakan dalam pengenalan suara berdasarkan domain frekuensi menggunakan skala Mel yang diadaptasi pada rentang frekuensi skala pendengaran manusia.

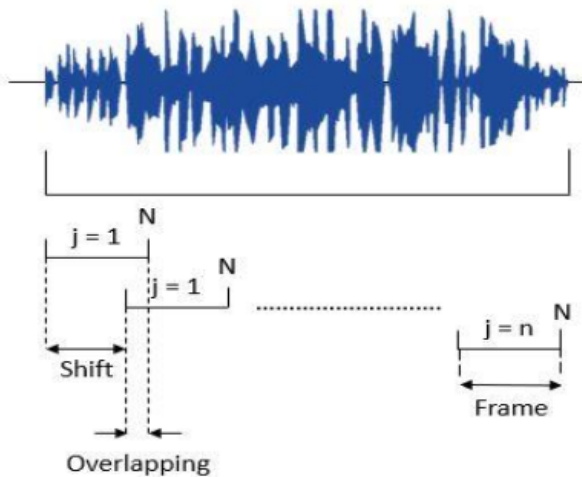
MFCC merupakan representasi dari *cepstral* nyata dari sinyal *short-time* yang telah di-*windowing* yang diturunkan dari transformasi *Fast Fourier Transform* (FFT) dari sinyal tersebut. Penyebab *cepstral* nyata ini berbeda adalah skala frekuensi non linier digunakan, sehingga dapat mendekati frekuensi pendengaran manusia. Selain itu, koefisien-koefisien ini kuat dan dapat diandalkan untuk berbagai macam variasi sesuai dengan kondisi pembicara dan perekaman [3].

2.4.1 Frame Blocking

Frame Blocking adalah proses pemotongan sinyal suara menjadi beberapa bagian sehingga memudahkan perhitungan dan analisa suara. Setiap potongan-potongan dari sinyal suara tersebut disebut *frame*. Setiap *frame* terdiri dari beberapa sampel tergantung waktu suara akan disampel dan berapa besar frekuensi *sampling*-nya [14].



Gambar 1. Proses Steganografi



Gambar 2. Frame Blocking pada MFCC [15]

Gambar 2 menunjukkan bahwa sinyal suara dibagi menjadi beberapa *frame* dan saling bertumpang tindih (*overlapping*). *Overlapping* dilakukan untuk memperlancar transisi dari *frame* ke *frame* sehingga menghindari adanya kehilangan ciri pada perbatasan perpotongan setiap *frame*. Panjang daerah *overlap* yang digunakan kurang lebih 30-50% dari panjang *frame*. *Frame* pada umumnya memiliki panjang antara 1030 ms [15].

2.4.2 Windowing

Frame blocking dapat menyebabkan terjadinya *aliasing* atau menyebabkan terjadinya diskontinuitas sinyal. Untuk mengurangi kemungkinan terjadinya efek-efek tersebut, maka setelah melalui proses *frame blocking* sinyal suara harus melewati proses *windowing*.

Windowing dilakukan pada setiap *frame* dengan tujuan untuk meminimalisir diskontinuitas antar dua *frame* yang berbatasan, khususnya pada bagian awal dan akhir [4]. Fungsi *windowing* yang paling sering digunakan dalam aplikasi *speech recognition* adalah *hamming* dan *hanning window*, dikarenakan keduanya memiliki karakteristik yang hampir sama.

2.4.3 Fast Fourier Transform (FFT)

FFT adalah salah satu metode untuk mengubah sinyal suara dari domain waktu menjadi domain frekuensi. Artinya, proses perekaman suara disimpan dalam bentuk digital berupa gelombang spektrum suara berbasis frekuensi yang kemudian menghasilkan pendeteksian gelombang frekuensi domain dalam bentuk diskrit [15]. FFT merupakan transformasi turunan dari *Discrete Fourier Transform (DFT)*, dimana jumlah perhitungan *digital* pada DFT dapat dikurangi secara signifikan. Sehingga, perhitungan *digital* terhadap spektrum-spektrum frekuensi pada FFT dapat diwujudkan secara sederhana dalam pengimplementasiannya.

2.4.4 Mel-Frequency Wrapping

Mel-Frequency Wrapping biasanya dilakukan menggunakan *filter bank*. *Filter bank* adalah salah satu bentuk dari filter yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari *frequency band* tertentu dalam sinyal suara. Pada MFCC, *filter bank* diterapkan dalam domain frekuensi.

Dalam penelitian [3], hubungan antara frekuensi bicara dengan skala Mel dapat ditetapkan seperti persamaan (1).

$$Frekuensi (Mel\ scaled) = \left(2595 \log \left(\frac{1 + f(Hz)}{700} \right) \right) \tag{1}$$

MFCC menggunakan *Mel-scale filter bank* dimana *filter* frekuensi yang lebih tinggi memiliki *bandwidth* yang lebih besar daripada frekuensi *filter* yang lebih rendah, tetapi resolusi temporal yang dimiliki keduanya sama [3].

2.4.5 Cepstrum

Cepstrum merupakan kebalikan dari *spectrum*. *Spectrum* merupakan kondisi ketika sinyal dianalisis dalam domain waktu, sedangkan *cepstrum* merupakan kondisi ketika sinyal dianalisis dalam domain frekuensi dan amplitudo dari sinyal tersebut diambil untuk menganalisis sinyal. *Cepstrum* biasa dipergunakan untuk mendapatkan suatu informasi dari sinyal suara yang diucapkan oleh manusia. Pada langkah terakhir ini, spektrum *log-mel (output dari filter bank)* dikonversi ke dalam domain waktu menjadi *cepstrum* menggunakan sebuah transformasi bernama *Discrete Cosine Transform (DCT)* [4]. DCT merentangkan koefisien-koefisien sesuai dengan signifikansi, dimana koefisien ke-0 dikecualikan karena sifatnya yang *unreliable* (tidak dapat diandalkan) [3]. Hasil dari proses inilah yang dinamakan dengan *Mel-Frequency Cepstrum Coefficients (MFCC)*.

2.5 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) merupakan metode yang terkenal untuk memperkirakan atau mengestimasi sub-ruang linier dengan sifat diskriminatif yang baik. LDA telah banyak digunakan baik dalam klasifikasi sinyal wicara maupun gambar [13]. Prinsipnya adalah untuk menemukan proyeksi data, dimana varians antara kelas-kelas itu lebih besar dibandingkan dengan varians di dalam kelas [17]. Sehingga, prinsip dasarnya adalah menemukan sebuah transformasi linear sehingga pengklasteran dapat dipisahkan setelah transformasi. Tujuan LDA adalah untuk menemukan transformasi optimal dengan cara meminimalkan jarak di dalam kelas (*within-class distance*) dan memaksimalkan jarak antar kelas (*between-class distance*) secara bersamaan, sehingga didapatkan diskriminasi yang maksimal.

2.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan suatu teknik yang dapat melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM masuk ke dalam kategori *supervised learning*, dimana dalam pengimplementasiannya memerlukan tahap *training* dan tahap *testing* [18]. SVM adalah metode lokal yang solusinya ditentukan secara eksklusif oleh vektor pendukung (*support vector*) sedangkan semua titik data lainnya tidak relevan dengan keputusan *hyperplane* [19]. Sehingga prinsip kerja SVM adalah menemukan fungsi pemisah (*hyperplane*) yang optimal, sehingga dapat memisahkan beberapa set data dari beberapa buah kelas yang berbeda pada *input space* secara maksimal.

3. Desain Model Sistem

Sistem steganalisis yang dirancang pada penelitian ini terdiri dari dua tahap, yaitu tahap pelatihan dan pengujian. Blok diagram perancangan model sistem pada data latih dan data uji ditunjukkan pada Gambar 3.

Pada tahap pelatihan, data sinyal wicara dimasukkan ke dalam perangkat, kemudian melalui proses MFCC dan LDA, yang keluarannya dijadikan data latih dan disimpan di dalam *database* latih. Sedangkan pada tahap pengujian, data uji dimasukkan ke dalam perangkat dan melalui proses yang sama seperti tahap pelatihan hingga proses LDA. Hanya saja setelah proses seleksi ciri, ciri latih yang telah disimpan di *database* saat proses pelatihan, digunakan sebagai ciri pembanding pada tahap klasifikasi. Adapun penjelasan dari setiap tahap pada konfigurasi sistem dapat diuraikan sebagai berikut:

1. Berkas sinyal wicara diekstraksi cirinya menggunakan MFCC, dimana di dalamnya terjadi transformasi FFT, yang berfungsi sebagai pengubah sinyal dari domain waktu ke domain frekuensi. Setelah melewati pengolahan, transformasi inverse FFT dilakukan, atau biasa

Tabel 1. Nilai SNR pada Speech Noise

Nama File	Nilai <i>Signal to Noise Ratio</i> (SNR)
Noise_AWGN-06	11,9232 dB
Noise_AWGN-07	13,5325 dB
Noise_AWGN-08	7,2717 dB
Noise_AWGN-09	17,1019 dB
Noise_AWGN-10	20,6797 dB
Noise_Echo-06	14,8303 dB
Noise_Echo-07	13,3948 dB
Noise_Echo-08	7,1612 dB
Noise_Echo-09	17,8911 dB
Noise_Echo-10	16,5837 dB
Noise_Gauss-06	30,8798 dB
Noise_Gauss-07	30,4278 dB
Noise_Gauss-08	28,9338 dB
Noise_Gauss-09	29,8252 dB
Noise_Gauss-10	29,0955 dB

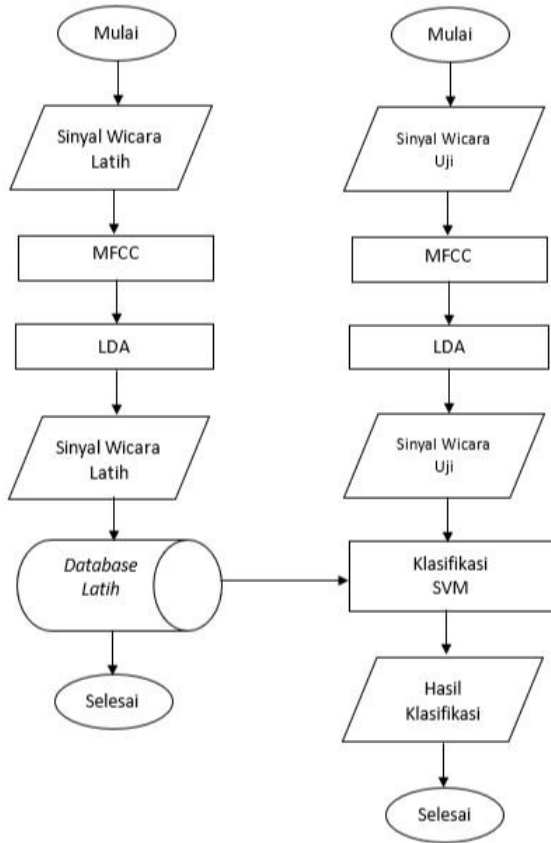
disebut DCT yang mengubah sinyal dari domain frekuensi menjadi domain waktu kembali. Output yang didapatkan dari proses ini berupa koefisien *Mel-Frequency Cepstrum*.

2. Hasil keluaran dari MFCC yang sudah berupa koefisien *Mel-Frequency Cepstrum* kemudian diseleksi cirinya menggunakan metode LDA. Tujuannya adalah untuk mengoptimalkan nilai diskriminan serta mempercepat proses pengujian.
3. Setelah dilakukan ekstraksi dan seleksi ciri, pengklasifikasian menggunakan metode SVM dilakukan, guna memisahkan beberapa set data dari kelas asli, *noise*, dan *stego* secara maksimal. Hal tersebut dilakukan dengan cara mencari *hyperplane* terbaik dari yang ada.
4. Jika hasil klasifikasi telah didapatkan, maka dapat dilakukan pengambilan keputusan dan proses steganalisis selesai.

4. Pengujian dan Analisis Sistem

Pengujian dilakukan menggunakan 15 buah data *speech* berformat .wav yang masing-masing berdurasi 10 detik. Data *speech* tersebut kemudian dikelompokkan menjadi 5 audio per kelompok untuk audio asli, penyisipan *stego*, dan penyisipan *noise*. Masing-masing data *speech* pada kelompok *stego* disisipi pesan rahasia dengan metode penyisipan LSB, DCT, dan DWT. Selain itu, masing-masing data *speech* pada kelompok *noise* juga disisipi tiga jenis *noise*, yaitu *gaussian*, *Additive White Gaussian Noise* (AWGN), dan *echo*. *Gaussian noise* merupakan *noise* yang mengikuti distribusi normal dan *white noise* adalah sinyal dengan energi yang sama pada semua frekuensi.

Biasanya akan sangat sulit menemukan contoh *white*



Gambar 3. Blok Diagram Perancangan Sistem Steganalisis

Tabel 2. Pengaruh Jenis Fungsi Windowing pada SVM Tahap 1

HAMMING		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		100%
HANNING		
No.	Kelas Audio	Akurasi
1	Asli	80%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		97,14%

noise yang tidak gaussian. Nilai Signal to Noise Ratio (SNR) pada setiap speech file uji yang tergolong dalam kelompok noise dapat dilihat pada Tabel 1.

Jumlah keseluruhan data data speech yang digunakan pada pengujian ini adalah 35, dengan spesifikasi 5 data asli, 15 data noise, dan 15 data stego. Klasifikasi SVM dilakukan melalui 2 tahap, pertama memisahkan dan mengklasifikasikan audio uji menjadi speech asli dan speech bukan asli dan kedua memisahkan dan

Tabel 3. Pengaruh Jenis Fungsi Windowing pada SVM Tahap 2

HAMMING		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	93,33%
3	Stego	66,67%
Akurasi Data Keseluruhan		82,86%
HANNING		
No.	Kelas Audio	Akurasi
1	Asli	80%
2	Noise	93,33%
3	Stego	73,33%
Akurasi Data Keseluruhan		82,86%

Tabel 4. Pengaruh Overlap pada SVM Tahap 1

OVERLAP		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		100%
NON-OVERLAP		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Bukan Asli	93,33%
Akurasi Data Keseluruhan		85,71%

Tabel 5. Pengaruh Overlap pada SVM Tahap 2

OVERLAP		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	93,33%
3	Stego	66,67%
Akurasi Data Keseluruhan		82,86%
NON-OVERLAP		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Noise	100%
3	Stego	33,33%
Akurasi Data Keseluruhan		62,86%

mengklasifikasikan audio uji yang tergolong bukan asli menjadi speech stego dan speech noise.

Performansi sistem dapat dilihat dari hasil akurasi sistem, yang mendeklarasikan bahwa setiap berkas sinyal wicara terdeteksi benar. Persamaan akurasi dapat ditulis dengan persamaan (2):

$$Akurasi = \frac{Jumlah\ data\ benar}{Jumlah\ seluruh\ data} \times 100\% \quad (2)$$

4.1 Analisis Pengaruh Jenis Fungsi Windowing Terhadap Akurasi

Menurut teori, fungsi *windowing* yang paling sering digunakan dalam aplikasi *speech recognition* adalah *hamming window* dan *hanning window*. Keduanya sama-sama memiliki fungsi untuk mengurangi terjadinya diskontinuitas pada sinyal setelah terjadi proses *frame blocking*. Maka dari itu, dilakukan penelitian untuk mengetahui fungsi *windowing* mana yang dapat menghasilkan akurasi yang lebih baik. Perbandingan hasil akurasi dapat dilihat pada Tabel 2 dan Tabel 3. Hasil akhir yang menjadi acuan pada penelitian ini dalam menentukan berhasil tidaknya sistem adalah hasil keluaran SVM tahap 2 yang sudah berupa apakah data audio yang diuji tergolong asli, *windowing*, atau *windowing*.

Dengan melihat hasil akurasi akhir kedua fungsi *windowing* dapat disimpulkan bahwa keduanya sama baik, dikarenakan sama-sama memiliki akurasi 82,86%.

4.2 Analisis Pengaruh Overlap Terhadap Akurasi

Menurut penjelasan pada penelitian [18], *overlapping* dilakukan untuk memperlancar transisi dari *frame* ke *frame*, sehingga menghindari adanya kehilangan ciri pada perbatasan perpotongan setiap *frame*. Maka dari itu, dilakukan studi untuk membuktikan tindakan mana yang dapat menghasilkan akurasi yang lebih baik. Dalam kondisi ini, *speech file* yang memiliki panjang 480.000 *samples di-overlap* dengan koreksi nilai M sebesar 100 *samples*, sehingga dalam satu audio terdapat 4786 *frames*, yang masing-masing *frame*-nya terdiri dari 1440 *samples*. Sementara, jumlah *frame* pada audio yang tidak melalui proses *overlap* adalah 334. Hasil akurasi pengaruh *overlap* pada SVM Tahap 1 dan Tahap 2, masing-masing dapat dilihat pada Tabel 4 dan Tabel 5 secara berturut-turut.

Dari hasil tersebut, dapat dilihat pengaruh ada tidaknya *overlapping* pada proses *frame blocking* terhadap akurasi sistem. Pada audio uji, steganalisis dengan menggunakan proses *overlap* memiliki akurasi yang lebih tinggi dibandingkan dengan yang tidak menggunakan proses *overlap*. Hal ini membuktikan bahwa *overlapping* pada sampel mampu mengurangi risiko adanya kehilangan ciri.

4.3 Analisis Pengaruh Panjang Frame Terhadap Akurasi

Dalam penelitian ini, dilakukan analisis pengaruh panjang *frame* terhadap akurasi sistem steganalisis yang didapatkan. Audio asli, serta audio yang telah disisipi *noise* dan pesan rahasia akan di-*framing* dengan tiga panjang *frame* yang berbeda yaitu 30 ms, 60 ms, dan 90 ms. Audio latihan dan audio uji kemudian dibandingkan dengan panjang *frame* tersebut untuk melihat

Tabel 6. Pengaruh Panjang Frame pada SVM Tahap 1

30 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		100%
60 ms		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		91,43%
90 ms		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		91,43%

performansi sistem. Hasil akurasi dapat dilihat pada Tabel 6 dan Tabel 7.

Dari hasil tersebut dapat dilihat bahwa semakin panjang suatu *frame*, maka semakin rendah akurasi yang didapat. Hal ini dapat terjadi dikarenakan adanya perbedaan banyak ciri *sampling*.

4.4 Analisis Pengaruh Jumlah Mel-Filter Bank Terhadap Akurasi

Menurut teori, jumlah *mel-filter bank* yang biasanya digunakan adalah berkisar pada angka 20-40. Maka dari itu, dilakukan penelitian untuk mengetahui berapa jumlah *mel-filter bank* yang dapat menghasilkan tingkat akurasi yang lebih baik. Hasil akurasi dapat dilihat pada Tabel 8 dan Tabel 9.

Dari hasil tersebut dapat kita lihat bahwa semakin banyak jumlah *mel-filter bank* yang dibangun pada tahap *mel-frequency wrapping*, maka semakin rendah tingkat akurasi yang didapat.

5. Kesimpulan

Berdasarkan hasil perancangan dan analisa sistem steganalisis dapat diambil beberapa kesimpulan sebagai berikut:

1. Sistem steganalisis menggunakan metode *Mel-Frequency Cepstral Coefficient* dan *Linear Discriminant Analysis* mampu membedakan audio asli, audio *noise*, dan audio *stego* secara baik meskipun belum sepenuhnya maksimal.
2. Dari keempat parameter yang diuji, hasil akurasi SVM tahap 1 (memisahkan audio asli dan bukan asli) selalu lebih baik dibandingkan dengan hasil akurasi setelah dilakukan SVM tahap 2 (memisahkan audio *stego* dan *noise*). Hal ini

Tabel 7. Pengaruh Panjang *Frame* pada SVM Tahap 2

30 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	93,33%
3	Stego	66,67%
Akurasi Data Keseluruhan		82,86%
60 ms		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Noise	93,33%
3	Stego	53,33%
Akurasi Data Keseluruhan		68,57%
90 ms		
No.	Kelas Audio	Akurasi
1	Asli	40%
2	Noise	93,33%
3	Stego	40%
Akurasi Data Keseluruhan		62,86%

Tabel 8. Pengaruh Jumlah *Mel-Filter Bank* pada SVM Tahap 1

30 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	100%
Akurasi Data Keseluruhan		100%
60 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	63,33%
Akurasi Data Keseluruhan		68,57%
90 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Bukan Asli	20%
Akurasi Data Keseluruhan		31,43%

dikarenakan karakteristik ciri audio *stego* dan *noise* yang lebih mirip dibandingkan karakteristik audio asli.

- Pada sistem steganalisis ini, perbedaan jenis fungsi *windowing* antara *hamming* dan *hanning window* tidak memiliki pengaruh yang terlalu signifikan, dikarenakan keduanya menghasilkan tingkat akurasi yang sama besar.
- Membagi panjang *frame* menjadi 30 ms dan melakukan *overlapping* sinyal pada tahap *frame blocking* dapat menghasilkan tingkat akurasi yang lebih baik dikarenakan terminimalisirnya risiko

Tabel 9. Pengaruh Jumlah *Mel-Filter Bank* pada SVM Tahap 2

30 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	93,33%
3	Stego	66,67%
Akurasi Data Keseluruhan		82,86%
60 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	60%
3	Stego	20%
Akurasi Data Keseluruhan		48,57%
90 ms		
No.	Kelas Audio	Akurasi
1	Asli	100%
2	Noise	33,33%
3	Stego	0%
Akurasi Data Keseluruhan		28,576%

kehilangan ciri pada saat pemotongan audio.

- Dari keseluruhan pengujian, performansi sistem terbaik yang didapatkan adalah 82,86%. Angka tersebut dicapai dengan kondisi audio latih dan audio uji menggunakan *hamming window*, terdapat *overlapping*, panjang *frame* 30 ms, serta jumlah *mel-filter bank* yang dibangun sebesar 20.

Daftar Pustaka

- N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *Computer*, vol. 31, no. 2, 1998.
- H. Liu, J. Liu, R. Hu, X. Yan, and S. Wan, "Adaptive audio steganography scheme based on wavelet packet energy," in *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2017 IEEE 3rd International Conference on*. IEEE, 2017, pp. 26–31.
- N. Dave, "Feature extraction methods lpc, plp and mfcc in speech recognition," *International journal for advance research in engineering and technology*, vol. 1, no. 6, pp. 1–4, 2013.
- S. Erani, B. Hidayat, and S. A. Wibowo, *Sistem Identifikasi Suara Pria dan Wanita Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient dan LPC Dengan Metode Klasifikasi SOM*, 2011.

- [5] J. Filipus, *Perbandingan Digital Steganografi Pada Media Image, Audio, Video dan Teks serta Kekuatannya Terhadap Steganalisis*, 2010.
- [6] Y. Kurniawan, "Studi metode steganalisis pada stegoimage."
- [7] R. Aeri, "Ekstraksi fitur menggunakan metode lda dan pemilihan eigen value pada cacat kertas duplek," *Jurnal Simantec*, vol. 3, no. 3, p. 142149, Dec 2013.
- [8] M. Z. Aminudin, B. Hidayat, and R. F. Umbara, *Steganalisis File Audio dengan Metode Statistik MFCC*, 2015.
- [9] A. Madan and D. Gupta, "Speech feature extraction and classification: A comparative review," *International Journal of computer applications*, vol. 90, no. 9, 2014.
- [10] M. Murugappan, N. Q. I. Baharuddin, and S. Jerritta, "Dwt and mfcc based human emotional speech classification using lda," in *Biomedical Engineering (ICoBE), 2012 International Conference on*. IEEE, 2012, pp. 203–206.
- [11] J. Pylkkönen, "Lda based feature estimation methods for lvcst," in *Ninth International Conference on Spoken Language Processing*, 2006.
- [12] B. Santosa, *Tutorial Support Vector Machines*, 2015.
- [13] I. B. Setiawan, T. A. S. Prasida, and M. Bezaleel, "Aplikasi noise reduction untuk perbaikan kualitas suara pada data audio menggunakan algoritma fastica," *Jurnal Teknologi Informasi*, vol. 8, no. 2.
- [14] E. M. Singla and M. H. Singh, "Paper on frequency based audio noise reduction using butterworth, chebyshev & elliptical filters," *Int. J. Recent Innov. Trends Comput. Commun*, vol. 3, pp. 5989–5995, 2015.
- [15] M. Singla and H. Singh, "Review paper on frequency based audio noise reduction using different filters," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 4, no. 5, p. 14061409, May 2015.
- [16] V. I. Sinisuka, "Steganalisis audio format wav menggunakan metode discrete wavelet transform dan linear discriminant analysis," 2017.
- [17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [18] A. Wicaksono, S. N. Endah, S. Adhy, and S. Sutikno, "Aplikasi speech recognition bahasa indonesia dengan metode mel-frequency cepstral coefficient dan linear vector quantization untuk pengendalian gerak robot," in *Prosiding Seminar Nasional Ilmu Komputer Undip 2014*.
- [19] T. Xiong and V. Cherkassky, "A combined svm and lda approach for classification," in *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 3. IEEE, 2005, pp. 1455–1459.