

DETEKSI STOMATITIS AFTOSA REKUREN BERBASIS PEMROSESAN SINYAL WICARA DENGAN METODA MFCC DAN DECISION TREE

STOMATITIS AFTOSA REKUREN DETECTION BASED ON SPEECH SIGNAL PROCESSING WITH MFCC AND DECISION TREE METHOD

Fadhilah¹, Dr. Ir. Bambang Hidayat, DEA.², Dr. Rudy Hartanto drg., MS., PAK.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom
¹anjirgabisa@students.telkomuniversity.ac.id,
²bhidayat@telkomuniversity.ac.id,
³juliahusny@hotmail.com

Abstrak

Stomatitis Aftosa Rekuren (SAR) yang pada masyarakat awam dikenal sebagai sariawan merupakan penyakit yang begitu sering terjadi, terkadang sampai menimbulkan masalah yang serius, hingga si penderita kesulitan berbicara, oleh karena itu untuk meningkatkan fasilitas dalam kedokteran gigi dan mulut maka penulis melakukan penelitian mengenai deteksi infeksi rongga mulut, dimana sistem ini berfungsi mendeteksi infeksi pada rongga mulut pasien, sehingga pasien tidak terlalu sulit menyampaikan keluhan yang dapat semakin memicu rasa sakit, dan melakukan pengukuran dengan beberapa parameter kalimat tertentu yang diucapkan pasien.

Tugas akhir (TA) ini melakukan deteksi infeksi pada rongga mulut pada sinyal wicara (*speech processing*). Proses deteksi dimulai dari input rekaman suara sehat dan sakit pada data latih dalam (*.wav) lalu dilakukan *preprocessing*. Sinyal suara ditransformasi MFCC dan ditambahkan 6 parameter statistik yang selanjutnya diproses oleh DT untuk menentukan apakah suara wicara tersebut terdeteksi infeksi atau tidak.

Hasil pengujian yang telah dilakukan adalah mengubah frekuensi sampling, koefisien MFCC dan ukuran *frame*. Akurasi terbaik didapat 92.85% pada frekuensi sampling sebesar 8000 Hz, koefisien MFCC sebanyak 20 dan ukuran *frame* sebanyak 512. Akurasi tersebut menggunakan 30 data latih dan data uji yang masing-masing terdiri dari 12 suara sakit dan 18 suara sehat.

Kata kunci : *Mel-frequency cepstral coefficients (MFCC), Decision Tree (DT).*

Abstract

Stomatitis Aftosa Rekuren (SAR), which in ordinary people known as canker sores is a very common disease, sometimes to cause serious problems, until the patient has difficulty speaking, therefore to upgrade the facilities in dentistry and mouth, writer wish to do research about detection of oral cavity infection, where the system is used to detect the patient's oral cavity, so patients do not complaints too difficult that can trigger the pain, and take measurements with some of the patient's specific sentence parameters.

This final assignment detects infections in the oral cavity in speech processing. The detection process starts from healthy and sick voice recording input on the training data in (*.Wav) then preprocessing is done. The sound signal is transformed by MFCC and 6 statistical parameters are added which are then processed by the DT to determine the speech sound is detected or not.

The test results that have been done are changing the sampling frequency, MFCC coefficient and frame size. The best accuracy is obtained 92.85% at the sampling frequency of 8000 Hz, the MFCC coefficient is 20 and the frame size is 512. The accuracy uses 30 training data and test data which each consists of 12 sick sounds and 18 healthy sounds.

Keywords : *Keywords: Mel-frequency cepstral coefficients (MFCC), Decision Tree (DT).*

1 Pendahuluan

Stomatitis aftosa rekuren (SAR) atau lebih dikenali oleh masyarakat awam dengan sariawan merupakan salah satu penyakit yang ulang kambuh pada mukosa mulut yang paling sering terjadi. SAR merupakan salah satu kasus yang sering dijumpai oleh dokter gigi diseluruh dunia sehingga dihasilkan beberapa penelitian-penelitian yang berhubungan dengan SAR [1].

Resiko terkena SAR cenderung meningkat pada kelompok sosioekonomi menengah ke atas, ini berhubungan dengan meningkatnya beban kerja yang dialami kalangan profesi atau jabatan-jabatan yang memerlukan tanggung jawab yang cukup besar, pada wanita dan individu yang stres, seperti mahasiswa yang sedang menghadapi ujian. Beberapa faktor yang berperan terhadap timbulnya *stomatitis aftosa rekuren* (SAR) antara lain trauma, infeksi bakteri atau virus dan gangguan sistem imun [2].

Tugas akhir (TA) ini melakukan perancangan deteksi infeksi pada rongga mulut pada sinyal wicara/*speech processing*. Proses deteksi berdasarkan fitur *Melfrequency cepstral coefficients* (MFCC) dan metode klasifikasi *decision tree* (DT). Sinyal wicara ditransformasi MFCC untuk mendapatkan fitur yang selanjutnya diproses oleh DT untuk menentukan apakah suara wicara tersebut terdeteksi infeksi atau tidak.

2 Dasar Teori

2.1 Sinyal Audio

Sinyal audio umumnya disebut sebagai sinyal yang dapat didengar oleh manusia. Energi yang terkandung dalam sinyal audio biasanya diukur dalam desibel. Sinyal audio biasanya berasal dari sumber suara yang bergetar dalam rentang frekuensi yang dapat didengar. Telinga manusia dapat mendengar bunyi antara 20 Hz hingga 20 KHz (20.000 Hz) sesuai batasan sinyal audio. Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 KHz merupakan frekuensi tertinggi yang dapat didengar. Nilai frekuensi di bawah 20 Hz disebut frekuensi infrasonik, sedangkan nilai frekuensi di atas 20 KHz disebut frekuensi ultrasonik [3].

2.2. Parameter Karakteristik Suara

Terdapat beberapa parameter untuk mengetahui karakteristik suara suatu sumber suara. Diantaranya yaitu frekuensi, intensitas dan amplitudo. Frekuensi adalah ukuran jumlah putaran ulang per peristiwa dalam selang waktu yang diberikan. Untuk memperhitungkan frekuensi, seseorang menetapkan jarak waktu, menghitung jumlah kejadian peristiwa, dan membagi hitungan ini dengan panjang jarak waktu. Hasil perhitungan ini dinyatakan dalam satuan hertz (Hz). Intensitas didefinisikan sebagai energi yang dipindahkan tiap satuan luas tiap satuan waktu. Karena energi tiap satuan waktu kita ketahui sebagai pengertian daya, maka intensitas bisa dikatakan juga daya tiap satuan luas. Amplitudo merupakan keras lemahnya bunyi atau tinggi rendahnya gelombang. Satuan amplitudo adalah desibel (db). Bunyi mulai dapat merusak telinga jika tingkat volumenya lebih besar dari 85 dB dan pada ukuran 130 dB akan mampu membuat hancur gendang telinga.

2.3 Preprocessing

Dalam *preprocessing*, sinyal suara melewati beberapa proses agar siap di ekstraksi ciri yang terkandung didalamnya. Tujuannya adalah untuk membuang informasi-informasi yang tidak diperlukan dalam ekstraksi ciri dan agar sinyal berada dalam kondisi yang sama dengan sinyal yang lain. Tahapan preprocessing meliputi *resampling*, *stereo to mono*, normalisasi amplitudo, *DC removal* dan *cropping*.

2.4 MFCC

Mel-Frequency Cepstral Coefficients (MFCC) dapat digunakan sebagai vektor ciri yang baik untuk merepresentasikan suara manusia dan sinyal musik. Lebih khusus lagi, MFCC telah terbukti bermanfaat untuk pengenalan suara [4].

Analisis suara pada *Mel-Frequency* didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu. Filter-filter tersebut memiliki jarak yang tidak seragam pada sumbu frekuensi, yaitu banyak filter pada daerah frekuensi rendah dan sedikit pada daerah frekuensi tinggi. Hal tersebut dikarenakan oleh sensitifitas pendengaran manusia yang tidak merata.

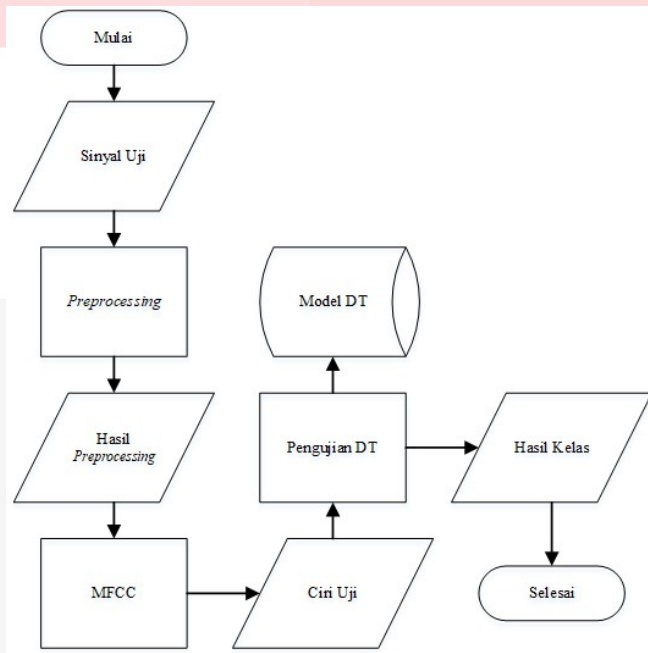
2.5 Decision Tree

Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah diinterpretasikan oleh manusia digunakan untuk pengenalan pola dan termasuk dalam pengenalan pola secara statistik. *Decision tree* berbentuk diagram alir seperti struktur pohon dan digunakan untuk mengklasifikasikan suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang sudah ada. Bentuk *decision tree* berdasarkan 3 tipe dari simpul

yaitu simpul *root*, simpul perantara, dan simpul *leaf*. Jalur pengujian data pertama melalui simpul *root* dan terakhir adalah melalui simpul *leaf* yang akan menyimpulkan prediksi kelas bagi data tersebut [5].

2.6 Skenario Pengujian

Berdasarkan tujuan pengujian maka skenario pengujian untuk menguji dan menganalisis performansi perangkat lunak untuk mendeteksi mengklasifikasi kualitas orang sakit maupun sehat dengan sinyal wicara adalah sebagai berikut sebanyak 60 subjek, dimana 24 subjek pertama adalah sakit dan 36 sisanya sehat, sample suara tersebut dibagi menjadi 30 data latih dan 30 data uji yang terdiri dari 12 suara sakit dan 18 suara sehat. Suara yang dijadikan data uji adalah suara yang berbeda dengan data latih. Hal ini berfungsi untuk menguji kemampuan sistem dalam mengenali suara sehat dan sakit.



Gambar 2.3 Flowchart Skenario Pengujian

3 Hasil dan Pembahasan

Pada sistem yang dibangun, skenario pengujian dibutuhkan terhadap pembacaan klasifikasi seseorang bahwa yang bersangkutan sehat atau sakit yang didapat saat pelatihan sistem. Dalam sistem ini digunakan 30 data sebagai data latih dan 30 data sebagai data uji. Masing-masing pengujian memiliki hasil yang direpresentasikan dalam bentuk tabel.

Tabel 4.1 Akurasi yang didapat dari perbandingan frekuensi *sampling*

Frekuensi <i>Sampling</i> (Hz)	Akurasi (%)
8000	92.8571
16000	85.7143
22050	90.4762
36000	90.4762
44100	95.2381
48000	97.6190

Pada tabel 4.1 akurasi di dapat dari frekuensi *sampling* yang berbeda, *frame* 512 dan koefisien 5, dapat dianalisis bahwa data tidak linier (tidak naik dan tidak turun) dan akurasi saat frekuensi *sampling* paling sedikit adalah 92.8571%, saat frekuensi *sampling* dinaikkan menjadi 16000 akurasi turun terdapat *bandwidth* sinyal frekuensi 16000 antara 4000 sampai 8000 yang merusak ciri. Saat menurunkan frekuensi *sampling* 36000 sampai 16000, untuk rentang *bandwidth* 18000 sampai 8000 akurasi turun, saat di redam akurasinya turun, frekuensi *sampling* menyebabkan peredaman sinyal. Frekuensi *sampling default* 44100 saat dinaikkan ke frekuensi *sampling* 48000 tidak ada informasi yang bertambah, hanya *sample* yang bertambah, kandungan informasinya tetap. Akurasi frekuensi *sampling* 48000 meningkat dari 95.2381% menjadi 97.6190% maka pada *bandwidth* antara 22050

sampai 24000 terdapat kandungan informasi yang mampu meningkatkan akurasi dan yang membuat ciri antara sehat dan sakit menjadi lebih baik.

Frekuensi *speech* berada pada rentang 300 hingga 3400, maka dengan menggunakan *bandwidth* sinyal 4000 atau frekuensi *sampling* 8000 proses pengujian sistem sudah tercukupi, dimana dengan frekuensi *sampling* 8000 sudah didapatkan akurasi yang tinggi sehingga tidak memerlukan frekuensi *sampling* yang besar karena *sample* yang terlalu banyak dan waktu yang terlalu lama.

Tabel 4.2 Akurasi yang didapat dari perbandingan ukuran *frame*

Frekuensi <i>Sampling</i> (Hz)	Ukuran <i>Frame</i>	Koefisien MFCC	Akurasi(%)
8000	512	5	92.85
8000	1024	5	95.23
8000	2048	5	95.23

Pada tabel 4.2 akurasi di dapat dari ukuran *frame* yang berbeda, frekuensi *sampling* dan koefisien yang sama, maka dapat dianalisis bahwa akurasi mengalami kenaikan sebesar 2.381% terdapat pada ukuran *frame* 1024 dan 2048.

Tabel 4.3 Akurasi yang didapat dari perbandingan dari koefisien MFCC dengan Ukuran *Frame* 512

Frekuensi <i>Sampling</i> (Hz)	Ukuran <i>Frame</i>	Koefisien MFCC	Akurasi(%)
8000	512	5	92.85
8000	512	10	92.85
8000	512	15	92.85
8000	512	20	92.85

Tabel 4.3 akurasi di dapat dari ukuran *frame* yang sama, frekuensi *sampling* sama dan koefisien yang berbeda, perbandingan koefisien MFCC pada ukuran *frame* 512 tidak berpengaruh perubahannya sehingga didapatkan akurasi dengan nilai yang sama.

Tabel 4.4 Akurasi yang didapat dari perbandingan dari koefisien MFCC, dengan Ukuran *Frame* 1024

Frekuensi <i>Sampling</i> (Hz)	Ukuran <i>Frame</i>	Koefisien MFCC	Akurasi(%)
8000	1024	5	95.23
8000	1024	10	92.82
8000	1024	15	88.09
8000	1024	20	92.85

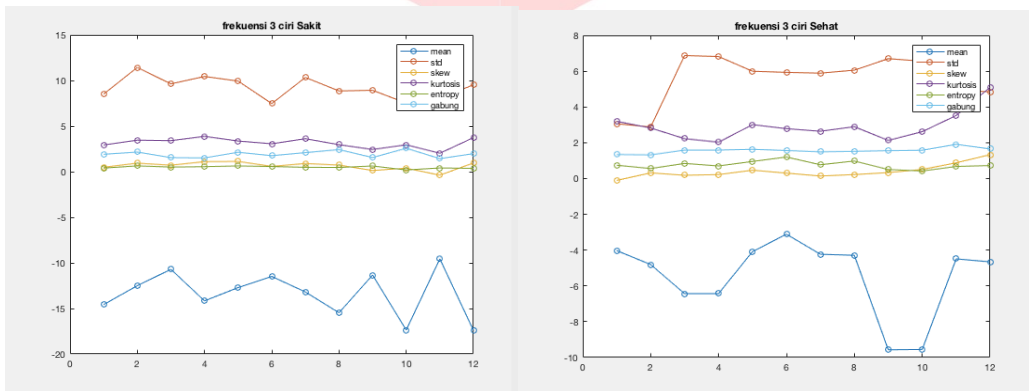
Tabel 4.4 akurasi di dapat dari ukuran *frame* yang sama, frekuensi *sampling* sama dan koefisien yang berbeda, akurasi yang di dapatkan tidak linier dimana akurasi tertinggi di dapat pada koefisien MFCC 5 dan akurasi terendah di dapat pada koefisien MFCC 15.

Tabel 4.5 Akurasi yang didapat dari perbandingan dari koefisien MFCC, dengan Ukuran *Frame* 2048

Frekuensi <i>Sampling</i> (Hz)	Ukuran <i>Frame</i>	Koefisien MFCC	Akurasi (%)
8000	2048	5	95.2381
8000	2048	10	95.2381
8000	2048	15	95.2381
8000	2048	20	85.7143

Tabel 4.5 akurasi di dapat dari ukuran *frame* yang sama, frekuensi *sampling* sama dan koefisien yang berbeda, menunjukkan bahwa akurasi terendah di dapatkan dari koefisien MFCC tertinggi, sedangkan pada koefisien 5,10,15 akurasi berada pada nilai yang sama.

Jenis ciri orde satu pada nilai frekuensi MFCC pada frekuensi *sampling* 8000, *frame* 512 dan koefisien 20 adalah sebagai berikut:



Gambar grafik 6 parameter statistik dalam kelas sehat dan sakit, dapat dianalisis bahwa 5 dari 6 parameter statistik menunjukkan garis yang stabil pada parameter standar deviasi, skewness, kurtosis, entropy dan gabungan, sesuai dengan hasil analisis frekuensi *sampling* 8000, ukuran *frame* 512 dan koefisien 20 dimana mendapat akurasi sebesar 92.8571%.

4 Simpulan dan Saran

Dari hasil pengujian dan analisis yang telah dilakukan terhadap perancangan sistem untuk mendeteksi sehat dan sakit seseorang maka dapat diambil beberapa kesimpulan, diantaranya adalah sebagai berikut:

1. Perancangan dan perealisasi sistem untuk mendeteksi Stomatitis Aftosa Rekuren melalui sinyal wicara dengan metoda MFCC dan klasifikasi *Decision Tree* terhadap data yang didapatkan berhasil dilakukan.
2. Hasil akurasi didapatkan paling baik dari perbandingan frekuensi *sampling* yaitu 92.85% pada ukuran *frame* 512 dan 8000 Hz frekuensi *sampling*. Dikarenakan suara yang dihasilkan saat ukuran *frame* 512 dan frekuensi *sampling* 8000 Hz merupakan suara ideal (tidak melengking tinggi dan panjang *frame* menyesuaikan fs).
3. Dengan frekuensi *sampling* 8000 Hz sudah didapatkan akurasi yang tinggi sehingga tidak memerlukan frekuensi *sampling* yang besar karena *sample* yang terlalu banyak dan waktu yang terlalu lama.
4. Klasifikasi *decision tree* di dapatkan melalui perhitungan *split point* pada grafik kelas sehat dan sakit, yang mana grafik tersebut mempresentasikan akurasi yang baik sesuai dengan pengujian frekuensi *sampling*, ukuran *frame* dan koefisien MFCC.

Saran untuk membantu pengembangan sistem selanjutnya adalah sebagai berikut.

1. Sistem dapat memperoleh data masukan secara *real time*, yaitu dapat langsung mendapatkan hasil kesehatan terhadap seseorang menggunakan kamera piranti, *voice recorder* atau *handphone*
2. Pengambilan sampel suara dilakukan diruangan kedap suara sehingga dapat mengurangi *noise*.
3. Diharapkan pada penelitian berikutnya dapat terkumpul variasi data yang lebih beragam dengan kualitas yang baik.
4. Diharapkan adanya pengembangan ke *platform* yang lain seperti *smart phone* menggunakan metode terbaik yang sudah terintegrasi ke *database*.

5 Daftar Pustaka

- [1] H. Fitri and D. Afriza, "Prevalensi stomatitis aftosa rekuren di panti asuhan kota padang," Padang.
- [2] S. Hudgens, K. Dineen, K. Webster, J. S. Lai, and D. Cella, "Assessing statistically and clinically meaningful construct deficiency/saturation: Recommended criteria for content coverage and item writing," in Rasch Measurement Transactions, 2000.
- [3] S. Ntalampiras and N. Fakotakis, "Speech/music discrimination based on discrete wavelet transform," 2008.
- [4] D. K. Pramesthi, "Sekilas tentang mel-frequency cepstral coefficient," 2011, (Date last accessed 22-Maret-2018). [Online]. Available: <http://nadytha-zildjian.blogspot.co.id/2011/07/sekilas-tentang-mel-frequency-cepstral.html>
- [5] Nandyka, "Decision tree? bukan pohon biasa!" 2008, (Date last accessed 22-Maret-2018). [Online]. Available: <https://nandyka.wordpress.com/2008/04/04/decision-tree-bukan-pohon-biasa/>