

Klasifikasi Suara Jantung Normal Dan Abnormal Menggunakan *Short-Time Fourier Transform* Dan *Convolutional Neural Network* Classification Normal And Abnormal Heart Sounds Using *Short-Time Fourier Transform* And *Convolutional Neural Network*

1st Muhammad Thariq Machaz
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
thariqmachaz@student.telkomuniversity.ac.id

2nd Purba Daru Kusuma
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
purbodaru@telkomuniversity.ac.id

3rd Achmad Rizal
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
achmadrizal@telkomuniversity.ac.id

Abstrak

Stetoskop adalah alat untuk mendiagnosa suatu penyakit untuk melakukan proses auskultasi. Auskultasi adalah proses memeriksa pasien dengan mendengarkan suara didalam tubuh yang bisa berupa suara jantung, paru-paru ataupun usus. Dari proses ini dapat diketahui suara didalam tubuh pasien itu normal ataupun abnormal. Pada penelitian ini diusulkan metode STFT dan deep learning CNN (Convolutional Neural Network) arsitektur Alexnet. Langkah analisis ini menggunakan short-time fourier transform dan kemudian data citra didapat dari STFT berupa citra plot sinyal, dilanjutkan ke proses klasifikasi data citra suara jantung normal dan abnormal dengan menggunakan metode CNN (Convolutional Neural Network). Dari uji klasifikasi CNN (Convolutional Neural Network) ini didapati tingkat akurasi dari proses pengujian ini menggunakan CNN Arsitektur Alexnet

dengan learning rate dan Iterasi/Epoch jumlah Terbaik yakni 0,00001 dan Jumlah iterasi 70 sehingga didapati akurasi suara jantung 91.07% untuk klasifikasi suara jantung normal dan abnormal menggunakan 56 data suara jantung. Presisi 88.46%, recall 92%, dan untuk f1 score 90.2%.

Kata kunci: Smartphone, Stethoscope, STFT, CNN, Heart Sounds, Signal, Image, Spectrogram, Alexnet

Abstract

Stethoscope is a tool for diagnosing a disease to perform the auscultation process. Auscultation is the process of examining the patient by listening to sounds in the body which can be sounds of the heart, lungs or intestines. From this process, it can be seen that the sound in the patient's body is normal or abnormal. This research proposes STFT method and deep learning CNN

(Convolutional Neural Network) Alexnet architecture. This analysis step uses a short-time Fourier transform and then image data is obtained from STFT in the form of signal plot images, followed by the classification process of normal and abnormal heart sound image data using the CNN (Convolutional Neural Network) method. From the CNN (Convolutional Neural Network) classification test, it was found that the level of accuracy of this testing process using the

CNN Alexnet Architecture with the best learning rate and Iteration/Epoch number was 0.00001 and the number of epochs was 70. Heart sound accuracy was found to be 91.07%. with 56 heart sound data, 88.46% precision, 92% recall, and 90.2% for f1 score.

Keywords: Smartphone, Stethoscope, STFT, CNN, Heart Sounds, Signal, Image, Spectrogram, Alexnet

I. PENDAHULUAN

Stetoskop adalah alat untuk mendiagnosa suatu penyakit yang berfungsi untuk melakukan proses auskultasi. Auskultasi merupakan suatu proses memeriksa pasien dengan mendengarkan suara didalam tubuhnya yang mana bisa suara jantung, paru-paru ataupun usus. Dari proses auskultasi ini dapat diketahui suara didalam tubuh pasien itu normal atau abnormal.

Penulis mengusulkan algoritma short-time fourier transform yang mana kemudian dapat membuat klasifikasi data suara jantung normal dan abnormal dengan CNN. STFT atau short-time fourier transform akan digunakan untuk merubah time domain ke time frequency domain sehingga audio suara jantung tersebut dapat ditampilkan dalam bentuk citra. Secara garis besarnya, CNN memanfaatkan proses konvolusi dengan menggerakkan sebuah kernel konvolusi (filter) berukuran tertentu ke sebuah gambar, komputer mendapatkan informasi representatif baru dari hasil perkalian bagian gambar tersebut dengan filter yang digunakan.

II. KAJIAN TEORI

A. Stetoskop Digital

Perkembangan zaman sudah sangat melesat dengan cepat, teknologi juga menjadi semakin kompleks. Di zaman serba teknologi pada sekarang ini, kita membutuhkan peralatan yang canggih, salah satunya pada bidang medis merupakan stetoskop. Stetoskop merupakan alat medis yang digunakan untuk melakukan auskultasi. Dan

stetoskop digital merupakan versi canggih dan lebih berkembang dari stetoskop biasa yang mana mampu merekam data audio kedalam bentuk file yang dapat diputar di Android maupun PC.



Gambar 1 Stetoskop Digital.
Sumber: www.stemoscope.com

B. Matlab

Matlab merupakan Bahasa pemrograman multi-paradigma dan komputasi numerik yang mana dikembangkan oleh Mathwork. Jutaan orang termasuk diantaranya adalah ilmuwan telah menggunakan Matlab sebagai Bahasa pemrograman mereka. Matlab dapat digunakan pada berbagai aplikasi di industry, akademik, *machine learning*, *deep learning*, *signal processing*, *communications*, *image processing*, *control system*, dan masih banyak lagi.

C. Short Time Fourier Transform

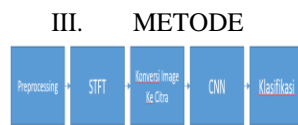
Short-time Fourier Transform (STFT), adalah metode transformasi *fourier* yang digunakan untuk menentukan frekuensi sinusoidal dan fase dari sinyal karena berubah dari waktu ke waktu [1]. Dalam proses prakteknya, untuk menghitung STFT adalah dengan membagi sinyal waktu yang lebih

panjang menjadi segmen-segmen yang lebih pendek dengan panjang yang sama dan kemudian menghitung transformasi Fourier-nya secara terpisah pada setiap segmen yang lebih pendek.

D. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu jenis metode yang biasa digunakan pada *image processing*. CNN bisa digunakan untuk mendeteksi dan mengenali object pada sebuah gambar maupun video. CNN adalah sebuah teknik yang terinspirasi dari cara mamalia maupun manusia, menghasilkan persepsi visual [2] [3].

Secara garis besarnya, CNN memanfaatkan proses konvolusi dengan menggerakkan sebuah kernel konvolusi (filter) berukuran tertentu ke sebuah gambar, komputer mendapatkan informasi representatif baru dari hasil perkalian bagian gambar tersebut dengan filter yang digunakan.



Gambar 1 Gambaran Sistem

Gambar diatas menjelaskan tentang bagaimana urutan sistem yang dibuat agar dapat dijalankan. Urutan sistem akan dilakukan sebagai berikut:

- a) Pada bagian ini, perlu file audio suara jantung. dijalankan dengan audio file dengan ekstensi .wav 44.1 khz yang diambil dari dataset yang telah disediakan dari stetoskop digital. Digunakan untuk membaca bentuk file audio tersebut dan mendapatkan informasi-informasi yang penting didalamnya.
- b) Selanjutnya pada bagian ini akan memprogram suara audio detakan jantung dan akan di STFT sehingga akan didapatkan beberapa nilai-nilai keluaran STFT (Short-time fourier

transform). Algoritma STFT akan menampilkan sinyal masukan dalam rentang waktu dan frequency menggunakan fungsi window.

- c) Berikutnya keluaran dari STFT berupa visual spektrogram dan akan mengambil gambar dari grafik ini yang mana fungsinya agar dapat dilanjutkan dengan metode klasifikasi CNN (Convolutional Neural Network) sehingga data suara jantung dapat diklasifikasi.
4. Kemudian setelah didapatkan gambar yang akan diklasifikasi, akan dicari nilai dari gambar tersebut sehingga memiliki ciri tersendiri dan didapatkan nilai suara jantung abnormal maupun normal.
5. Klasifikasi data merupakan bagian akhir dari analisis, setelah didapatkan nilai ciri khas dari gambar yang telah diekstraksi, kemudian dari nilai itu akan dibuat klasifikasi sehingga akhirnya program dapat membedakan suara jantung normal dengan yang abnormal.

A. Perancangan Program



Gambar 2 Diagram AliH Klasifikasi Suara Jantung

Diagram diatas menunjukkan alur dari klasifikasi suara jantung menggunakan STFT dan kemudian dilanjutkan dengan proses Convolutional Neural Network (CNN).

B. Preprocessing

Pada bagian ini dilakukan proses normalisasi suara jantung dan *remove data*

suara jantung yang berbeda akibat proses perekaman suara jantung yang berbeda-beda. Data tersebut dapat ditulis sebagai berikut:

$$S = S - \text{mean}(S); S = S/\text{max}(S)$$

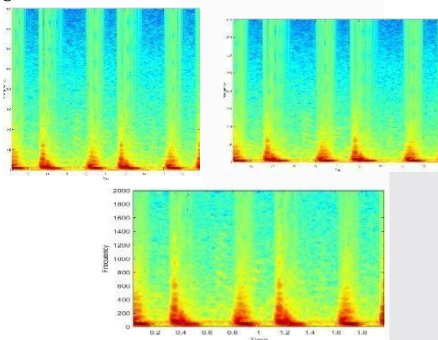
Yang mana S merupakan *audio sample* dengan *fs frequency sampling* (hz).

C. Short-Time Fourier Transform adalah tahap untuk menganalisa frekuensi sinyal dan menjadikan sinyalnya tersebut tersegmentasi pada interval waktu tertentu. Sinyal tersegmentasi ditransformasikan menggunakan Fast Fourier Transform (FFT)

ke domain frekuensi. Dengan STFT akan menampilkan sinyal masukkan dengan rentang waktu dan frequency menggunakan *window function*. Perhitungan STFT secara matematis dapat ditulis sebagai berikut :

$$XSTFT [m, n] = \sum_{k=-L}^L x[k]g[k - m]e^{-j2\pi nk/L}$$

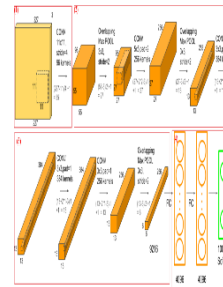
Yang mana x(k) merupakan sinyalnya dan g(k) merupakan L-point window. jadi STFT bisa dikatakan sebagai fourier transform dari sinyal x(k) yang mana telah menggunakan fungsi window g(k). Parameter yang digunakan dalam proses STFT sebagai media klasifikasi ialah sebagai berikut ini: Window function Keiser, Window dan overlap length: 25-20, 200-100, dan 500-475. FFT 512 dan 256



Hasil dari tiap varian sama tetapi dari STFT suara jantung normal 25-20-512 FFT dengan STFT suara jantung normal 500-475-512 FFT itu akurasi sama dengan K=5 dan d=1. Karena perbedaan N-point FFT mempengaruhi resolusi fitur yang akan dihasilkan, oleh suara jantung, dapat diambil

kesimpulan pemilihan window length dan overlaps akan mempengaruhi resolusi waktu.

D. Proses Klasifikasi CNN (Alexnet)



Gambar 3 Arsitektur Alexnet

Dalam arsitektur *alexnet*, citra akan melewati beberapa layer, yaitu 5 layer konvolusi dan 3 layer max pooling sehingga citra dapat dikenal. Dalam kasus ini akan menggunakan citra plot STFT rgb yang digunakan untuk klasifikasi suara jantung.

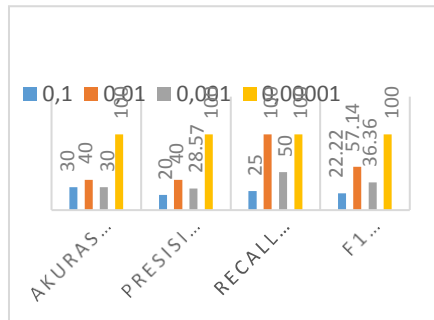
IV. HASIL DAN PEMBAHASAN

Dilakukan 3 tahapan, yang mana tahap pertama yaitu pengujian dengan 3 parameter STFT yang berbeda. Tahap kedua dilakukan dengan jumlah *learning rate* yang berbeda, dan tahap yang ketiga melakukan pengujian dengan nilai *epoch* yang berbeda.

A. Pengujian Dengan Varian Parameter STFT

Pengujian ini dilakukan untuk mencari nilai akurat paling tinggi menggunakan 3 parameter STFT kaiser yang berbeda. Yaitu kaiser dengan parameter: (S,kaiser (25,5),20,512,Fs) yang akan dijadikan sebagai data A(S,kaiser(200,5),100,256,Fs) yang akan dijadikan sebagai data B, Dan

(S,kaiser(500,5),475,512,Fs) yang akan dijadikan sebagai data C

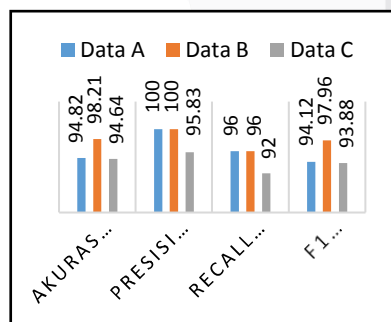


Gambar 4 Pengujian Dengan Varian Parameter STFT

Hasil dari pengujian pertama didapatkan hasil yang berbeda antar data. Pada gambar menunjukkan data B sebagai parameter citra STFT yang memiliki nilai akurasi paling tinggi, yaitu sebesar 98.21%, presisi 100%, recall 96%, dan nilai f1 score 97,96%.

B. Pengujian Mencari Nilai Learning Rate Terbaik

Pengujian ini dilakukan untuk mencari nilai learning rate terbaik dengan parameter STFT data B yang didapati akurasi paling tinggi pada pengujian tahapan pertama. Nilai learning rate yang akan dilakukan pengujian untuk mencari nilai yang terbaik ialah 0,1, 0,01, 0,001, dan 0,00001. Untuk parameter lainnya menggunakan iterasi epoch 70, batch size 64 dan citra dari parameter STFT data B yang telah dilakukan pada pengujian sebelumnya. Hasil pengujian ini dapat dilihat pada gambar berikut:

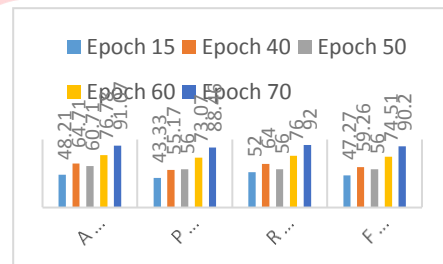


Gambar 5 Pengujian Mencari Nilai Learning Rate Terbaik

Hasil dari pengujian mencari nilai learning rate terbaik pada data B dari parameter STFT pengujian sebelumnya. Dapat dilihat dari Gambar 4.2, menunjukkan hasil pengujian nilai learning rate terbaik ada pada nilai 0,00001 yang memiliki nilai akurasi paling tinggi yaitu mencapai 100%. Dengan presisi 100%, recall 100%, dan f1 score juga mencapai 100%.

C. Pengujian Mencari Epoch Terbaik

Tahapan ketiga dilakukan proses pengujian mencari nilai epoch yang terbaik dengan menggunakan parameter stft citra data B, learning rate dengan nilai 0,00001 yang merupakan learning rate terbaik dari tahapan sebelumnya. Menggunakan batch size 64. Nilai epoch yang akan diuji ialah dengan nilai 15, 40, 50, 60, dan 70. Hasil dari pengujian ini dapat dilihat dari gambar berikut



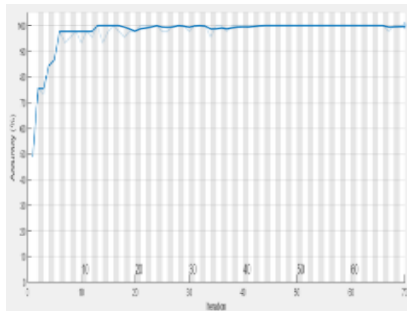
Gambar 6 Pengujian Epoch Terbaik

Pada Gambar 6 dapat dilihat bahwa hasil dari pengujian mencari nilai epoch terbaik menggunakan parameter STFT data B mengalami kenaikan pada tiap jumlahnya, Akurasi epoch tertinggi didapati pada nilai 70 yang mendapat tingkat akurasi 91.07%, presisi 88.48%. recall 92%, dan f1 score 90.2%.

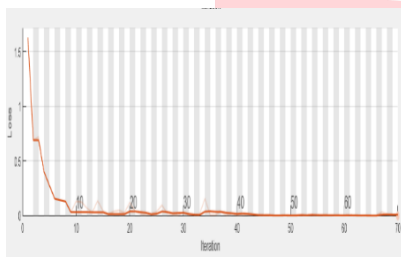
D. Pengujian Skenario Terbaik

Pada tahapan ini, akan dilakukan pengujian dengan scenario terbaik berdasarkan pengujian sebelumnya. Yaitu dengan epoch/iteration berjumlah 70, learning rate 0,00001, dan parameter STFT data B dengan citra sinyal STFT kaiser point FFT 254, overlap 100, dan window length 200. Berikut hasil dari pengujian ini dapat

dilihat grafik akurasi dan *loss* dapat dilihat dari Gambar 4.4 dan Gambar 4.5.



Gambar 7 Data Loss



Gambar 7 Data Train Accuracy

Dengan pengujian ini dapat dilihat dari Tabel 1 bahwa dari 56 data suara jantung yang terdeteksi benar ada 51 data, dan 5 dinyatakan salah. Terdapat 25 suara jantung normal yang mana akurasi mendapat 2 abnormal dan dari 31 suara jantung abnormal mendapatkan 3 normal. Dari pengujian ini kita akan langsung memulai tahapan berikutnya dengan *confussion matrix*.

E. Confusion Matrix

Dari *confusion matrix* ini akan memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang benar. *Confusion matrix* bisa berbentuk tabel yang berisi nilai dari suatu model klasifikasi pada data yang telah diuji. Berdasarkan pengujian scenario terbaik, dapat kita gambarkan dengan gambar berikut ini

		True	
		Suara Jantung Normal	Suara Jantung Abnormal
Predicted	Prediksi Normal	23	3
	Prediksi Abnormal	2	28

Gambar 8 Confusion Matrix

Dilihat dari Gambar 4.6 terdapat 2 kelas yang mana ada suara jantung normal dan abnormal. Dari total 56 dikali dengan data suara jantung nilai *true positif* dan ditambah *true negative* dan kemudian dibagi total *false negative* ditambah *false positif*, dan ditambah total tambah *true positif* dan *true negative* dan dikali 100% maka didapat nilai akurasi sebesar 91.07%, presisi 88.46%, recall 92%, dan f1 score 90,2%.

V. KESIMPULAN

Dari analisis pada penelitian ini dapat ditarik kesimpulan berupa File audio dari file berekstensi .wav frekuensi 44.1khz dapat dipresentasikan menjadi bentuk citra berupa *spectrogram* yang memiliki waktu 2 detik. Dengan metode STFT file audio dapat dipresentasikan menjadi bentuk plot segmen sinyal yang sederhana sehingga mudah untuk dianalisis. CNN dengan arsitektur Alexnet dengan awalan Pada input image berukuran 227 x 227, yang kemudian dilakukan convolusi dengan stride 4 dan 96 kernels dengan ukuran 11 x 11 sehingga pada hasil akhir menjadi ukuran 6 x 6. Pada proses CNN Alexnet ada 5 proses layer konvolusi dengan 3 layer *max pooling*. Nilai akurasi yang berhasil didapat dari 25 data suara jantung normal dan 31 suara jantung abnormal didapati dengan pengujian scenario terbaik dan dengan *confusion matrix* didapatkan akurasi 91,07%, presisi 88.46%, recall 92%, dan f1 score 90,2%. Dan *learning rate* dengan akurasi tertinggi didapatkan oleh nilai 0,00001 dan *epoch* yang digunakan sebanyak 70 yang memiliki akurasi yang baik. Berdasarkan pengujian dan analisis yang telah dirancang diatas, dapat diperoleh beberapa saran untuk pengembangan selanjutnya yaitu, Untuk data training

berikutnya cukup lakukan 1 kali saja proses training agar membuat waktu semakin efektif dan efisien untuk melakukan proses klasifikasi suara jantung normal dan abnormal. Mengumpulkan lebih banyak dataset agar membuat penelitian semakin baik dan membuat tingkat akurasi data semakin lebar.

REFERENSI

- [1] Y. Kalfika, A. Rizal and B. Prasetya, Analisis Kinerja Algoritma Short Time Fourier Transform (STFT) Untuk Deteksi Sinyal Carrier

Frequency Hopping Spread Spectrum (FHSS) CDMA, 2008.

- [2] Xing; , Xiao; , Yongping; , Chuangbai; , Wu; , Yifei; , Ding; , Ziming, A Convolutional Neural Network for Aspect Sentiment Classification, 2018.

- [3] J. Tanida, W. Zhang, I. Kazuyoshi and I. Yoshiki, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," vol. 29, pp. 29-32, 1990.

