Penerapan *Machine Learning* Pada Deteksi *Watermark* Berbasis *Multi Bit* Ss Yang Tidak Sempurna Akibat Serangan

1st Muhammad Giffary
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
mgffry@student.telkomuniversity.ac.id

2nd Gelar Budiman
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
gelarbudiman@telkomuniversity.ac.id

3rd Khaerudin Saleh
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
khaerudin@telkomuniversity.ac.id

***Abstrak* — Kemajuan teknologi dan penggunaan internet yang berkembang pesat saat ini memudahkan setiap orang untuk mendapatkan dan membagikan informasi dalam bentuk media digital. Berbagai data seperti gambar dan suara dapat dengan mudah disebarluarkan dan diakses oleh berbagai kalangan. Hal tersebut menjadikan hak kepemilikan suatu informasi atau yang biasa disebut dengan hak cipta tidak lagi terjamin. Oleh karena itu diperlukan sebuah metode untuk verifikasi dan autentifikasi data digital untuk melindungi hak cipta. Pada kasus ini *watermarking* dapat menjadi solusi permasalahan. Penelitian ini menerapkan *Machine Learning* dalam *audio watermarking* berbasis *Multi-bit* SS untuk mendeteksi watermark yang tidak sempurna akibat serangan.. *Watermark* diekstraksi dari *watermarked audio* dan kemudian dideteksi menggunakan *Convolutional Neural Network* (CNN) dengan parameter akurasi teks. Berdasarkan hasil simulasi, nilai rata-rata *Signal to Noise Ratio* (SNR) yang diperoleh sebesar 16.03, nilai rata-rata *Objective Difference Grade* (ODG) yang diperoleh sebesar -1.52, nilai *capacity* (C) yang diperoleh sebesar 57.42, nilai rata-rata *Bit Error Rate* (BER) yang diperoleh sebesar 0.26, dan nilai rata-rata akurasi teks yang diperoleh sebesar 30.72%.**

***Kata Kunci:* Audio watermarking, Multi-bit Spread Spectrum, Machine Learning, Convolutional Neural Network**

I. PENDAHULUAN

Kemajuan teknologi dan penggunaan internet yang berkembang pesat saat ini memudahkan setiap orang untuk mendapatkan dan membagikan informasi dalam bentuk media digital. Berbagai data seperti gambar dan suara dapat dengan mudah disebarluaskan dan diakses oleh berbagai kalangan. Hal tersebut menjadikan hak kepemilikan suatu informasi atau yang biasa disebut dengan hak cipta tidak lagi terjamin. Maka diperlukan sebuah metode untuk verifikasi dan autentifikasi data digital untuk melindungi hak cipta. Pada kasus ini *watermarking* dapat menjadi solusi permasalahan.

Watermarking adalah salah satu cabang dari ilmu steganografi. Steganografi sendiri merupakan ilmu yang mempelajari metode untuk menyembunyikan suatu informasi yang bersifat rahasia ke dalam media informasi lainnya. *Watermarking* merupakan proses penyisipan watermark dalam suatu *host file* sebagai informasi kepemilikan tanpa merubah file aslinya yang bertujuan untuk melindungi orisinalitas dari sebuah karya . *Watermarking* tediri dari dua tahap utama, yaitu proses penyisipan dan proses ekstraksi. Dimana pada proses penyisipan watermark disisipkan ke dalam *host file*, sedangkan dalam proses ekstraksi *watermark* yang telah disisipkan akan dideteksi. *Watermark* harus memiliki sifat tidak boleh tampak, tidak terdeteksi serta tahan terhadap serangan. Ketiga sifat tersebut biasa disebut *imperceptibility, security,* dan *robustness*. Pada penelitian ini akan digunakan metode berbasis *Multi-bit Spread Spectrum* (*Multi-bit* SS) dan *Convolutional Neural Network* (CNN). Pada proses penyisipan, digunakan *Pseudo Noise Code* (PN *Code*) untuk penyebaran data agar *attacker* yang melakukan penyerangan terhadap *watermark* sulit mengidentifikasi posisi *watermark*, sehingga data tidak dapat diekstraksi tanpa mengetahui PN *code* nya [1]. *Convolutional Neural Network* (CNN) bisa dilatih menggunakan data *training*, sehingga CNN dapat digunakan sebagai pendeteksi *watermark* yang tidak sempurna akibat serangan agar bisa dikembalikan menjadi *watermark* sebelum terkena serangan.

Perancangan sistem *audio watermarking* dapat dilakukan dengan menggunakan berbagai metode. Pada penelitian [2] menggunakan metode Spread Spectrum (SS) dan *Discrete Cosine Transform* (DCT). Metode SS digunakan untuk meningkatkan ketahanan terhadap serangan dan DCT memiliki fungsi untuk mengubah sinyal dari domain waktu ke domain frekuensi. Penelitian ini menggunakan *detection rate* (DR) untuk mengukur *robustness* dari metode yang digunakan. Hasil dari penelitian [2] ini memiliki nilai DR yang tinggi sehingga metode yang digunakan kuat terhadap serangan yang berbeda. Pada penelitian [3] menjelaskan bahwa *Discrete Wavelet Transform* (DWT) dan *Direct Sequence Spread Spectrum (DSSS)* digunakan agar *watermark* dapat disisipkan di frekuensi rendah yang kurang terdengar. Hasil penelitian [3] ini memperoleh nilai SNR (22.64), ODG (-0.22), dan Payload (83.00).

Pada penelitian [4], menjelaskan *data hiding* menggunakan metode *Compressive Sampling* (CS) dengan *Multi-bit* SS untuk mengambil sampel sinyal audio dan menyisipkan *watermark* pada saat yang sama sehingga ukuran sinyal sampel nya lebih kecil. Hasil penelitian [4] ini memiliki *imperceptibility* yang tinggi dengan nilai *payload* (729-5292) bps, *Compression Ratio* (CR) (1.47-4.84). Sedangkan untuk nilai ODG nya berada di rentang (-0.94 -0.74) dan nilai BER tertingginya 13%.

Pada penelitian [5], menjelaskan *Convolutional Neural Network* (CNN) untuk mendeteksi beberapa jenis serangan pada *image watermarking*. Model dari CNN yang digunakan berupa beberapa layer yang berisi *Input, Convolutional layer, Pooling layer, Fully connected layer,* dan *Softmax classifier*. Parameter dari model CNN tersebut menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU), *optimizer* Adam, 20 epoch, dan 50 *batch size.* Untuk pengukuran performa menggunakan PSNR, NC, dan SSIM. Hasil dari penelitian [5] mendapatkan performa yang baik akurasi 98% untuk mendeteksi beberapa jenis serangan pada *watermark* yang sudah di ekstraksi.

Pada penelitian [6], menjelaskan *Deep Learning-based audio-in-image watermarking* dengan menggunakan *Similarity Network* yang digunakan untuk mendeteksi *watermark* yang terganggu. Untuk 100 epoch, nilai akurasi yang didapatkan dari *Similarity Network* setelah di training adalah 99.44%. Nilai rata-rata *Root Mean Square Error* (RMSE) adalah 0.009452 berdasarkan lebih dari 5.800 *audio watermark* yang berbeda.

Pada penelitian [7], menjelaskan *digital image watermarking* meggunakan *Convolutional Neural Network* (CNN). Network pada proses penyisipan menggunakan CNN dengan

mempertahankan resolusi untuk menampilkan gambar yang diberi *watermark*. Network pada proses ekstraksi juga menggunakan CNN dengan mengurangi resolusi untuk menampilkan informasi *watermark*. Nilai rata-rata PSNR yang didapat adalah 40.58 dB. Untuk ketahanan terhadap serangan, ketika diberi serangan yang merubah nilai piksel didapatkan nilai BER yang baik yaitu kurang dari 10%.

II. DASAR TEORI

A. *Machine Learning*

Pada tahun 1970-an, sistem pakar banyak digunakan dibanyak sektor untuk membuat perangkat lunak pembuat keputusan (*Decision Making System*). Sistem pakar ini dikembangkan dengan memberi beberapa masukkan ke dalam sistem dan memberikan jawaban berdasarkan masukkan pola yang diberi pada saat sistem tersebut akan dibuat. Hal ini yang dapat menjadi masalah ketika data tersebut bertambah tetapi pola yang diberikan tidak begitu mendukung. Sistem pakar juga memiliki kelemahan ketika adanya kasus yang tidak diprediksi sebelumnya.

Dengan pemanfaatan perangkat lunak, komputer memiliki kemampuan yang luas. Hasil komputasi juga umumnya memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan pendekatan konvensional yang dilakukan manusia. Namun, terdapat situasi-situasi tertentu di mana keakuratan komputer tidak dapat mencukupi. Sebagai contoh, dalam konteks penyaringan *email spam*, tugas ini sulit diatasi sepenuhnya oleh pendekatan konvensional. Salah satu cara untuk mengatasi tantangan ini adalah dengan membangun kumpulan data yang memadai, yang kemudian dapat digunakan untuk mengembangkan model *Machine Learning* guna melakukan *filtering* secara otomatis.

B. Audio

Audio diartikan sebagai suara atau reproduksi suara. Audio merupakan getaran suatu benda yang menghasilkan sebuah bunyi atau suara, sehingga suara tersebut dapat didengar oleh telinga manusia. Satu-satunya tempat dimana suara tidak dapat merambat adalah ruangan hampa udara. Untuk bisa ditangkap oleh telinga manusia getaran tersebut harus berkekuatan 20 Hz hingga 20kHz sesuai batasan sinyal audio. Karena pada dasarnya sinyal audio adalah sinyal yang dapat diterima oleh telinga manusia. Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 kHz merupakan frekuensi tertinggi yang dapat didengar. *Sample rate* pada audio adalah banyaknya suara atau getaran yang direkam dalam satu detik dengan satuan *Hertz* (Hz). *Sample rate* yang biasa digunakan adalah 44100 Hz. Audio yang sama dengan *sample rate* berbeda menghasilkan kualitas audio yang berbeda pula.

Audio memiliki beberapa format, seperti:

1. *Waveform Audio* (WAV)
   Format WAV merupakan *file* audio yang tidak terkompres, jadi seluruh sample audio disimpan dalam bentuk digital.
2. *Advanced Audio Coding* (AAC)
   Format AAC merupakan bagian dari MPEG. Suara yang dihasilkan lebih bagus walaupun dalam bit rendah.
3. *Windows Media Audio* (WMA)
   Format WMA adalah format khusus untuk Windows Media Player yang ada pada Windows.

C. *Watermarking*

Watermarking adalah salah satu bidang studi yang fokus pada teknik penyisipan informasi. Teknik ini juga sering disebut sebagai teknik penyembunyian data. Meskipun begitu, perbedaan mendasar terdapat antara *watermarking* dan steganografi. *Watermarking* mengandalkan keterbatasan sistem sensor manusia seperti penglihatan dan pendengaran. Dengan memanfaatkan keterbatasan ini, metode *watermarking* dapat diterapkan pada berbagai jenis media digital.

Watermarking merupakan proses penyisipan data atau *watermark* ke dalam suatu elemen multimedia seperti citra, audio, dan video. Data yang disisipkan harus dapat diekstrak kembali, maka dari itu *watermarking* memiliki dua proses utama yaitu penyisipan data dan ekstraksi data [9].

D. *Audio Watermarking*

Audio watermarking merupakan suatu proses penyisipan data digital sebagai sebuah *watermark* ke dalam *host file* yang berupa audio. *Watermarking* pada sinyal audio mempunyai tantangan yang lebih dibandingkan dengan *watermarking* pada citra atau video. *Audio watermarking* memanfaatkan ketidaksempurnaan sistem pendengaran manusia. *Watermark* sendiri mengandung hak cipta untuk melindungi *file* audio. Perancangan sistem *audio watermarking* harus memenuhi tiga hal berikut ini [10]:

1. *Imperceptibility*
   Kualitas suatu *audio watermark* harus sama dengan audio originalnya, sehingga *watermark* tidak dapat terdeteksi oleh indera pendengaran manusia.
2. *Robustness*
   Watermark harus tahan terhadap berbagai serangan digital dan tidak berubah oleh berbagai transformasi yang terjadi.
3. *Capacity*
   Capacity merepresentasikan jumlah bit yang dapat disisipkan pada audio *host*. Semakin tinggi *capacity* maka akan semakin banyak pula bit yang dapat disisipkan.

Berdasarkan domain di mana mereka disisipkan, teknik *watermarking* pada audio diklasifikasikan menjadi dua kategori utama, yaitu *watermarking* temporal dan *watermarking* spektral. *Watermarking* temporal adalah melakukan penyisipan dalam domain waktu pada audio *host*, sementara pada *watermarking* spectral, terlebih dahulu dilakukan transformasi dari domain waktu ke domain frekuensi sebelum penyisipan *watermark* dilakukan pada komponen frekuensi.

E. Metode Penelitian

1. *Multi-bit Spread Spectrum* (*Multi-bit SS*)

Multi-bit Spread Spectrum merupakan pengembangan dari metode *Spread Spectrum* konvensional. Dengan menggunakan *Multi-bit* SS, bit-bit *watermark* akan disebar melalui *spectrum* dari sinyal *host*. Keuntungan yang diperoleh dari proses penyebaran ini adalah didapatkan *robustness* dan *security* yang kuat dikarenakan setiap bit yang disebar adalah bit-bit yang kecil dan sulit dideteksi. Pengembangan dilakukan dengan meningkatkan kapasitas *watermark* yang lebih baik menggunakan *pseudo noise sequence* untuk menyisipkan beberapa *bit watermark*. Pemilihan PN *Sequence* dilakukan dengan mengkonversi bilangan biner kedalam decimal. Untuk proses perhitungannya menggunakan rumus sebagai berikut [10]:

$$Y = X + \alpha W \quad (1)$$

dimana:
Y = sinyal hasil *watermarking*
X = sinyal *audio host* yang sudah ditransformasi
α = faktor penguat *watermark*
W = bit-bit *watermark*

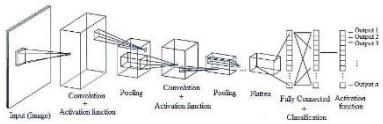
Pemilihan PN *Sequence* dilakukan dengan mengkonversi bilangan biner ke dalam desimal. Sehingga rumus penyisipan menjadi sebagai berikut:

$$Y(n) = X(n) + P_t \quad (2)$$

dengan P_t merupakan PN *Sequence*, $Y(n)$ merupakan sinyal watermarked audio, dan $X(n)$ sinyal audio host asli.

2. *Convolutional Neural Network* (CNN)

Convolutional Neural Network adalah salah satu jenis neural network yang biasa digunakan pada data image. CNN merupakan arsitektur yang mengimplementasikan *deep learning*. 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. CNN cocok digunakan untuk berbagai tugas pencitraan seperti pengenalan objek, deteksi objek, dan segmentasi. CNN dapat dilatih dengan menggunakan data *training*.



GAMBAR 2. 1
Convolution Neural Network

Convolution layer adalah lapisan pertama yang menjadi bagian utama dari CNN. *Convolution layer* terdiri dari neuron yang membentuk sebuah *filter* dengan area kecil (piksel) berupa panjang dan tinggi. Suatu piksel merupakan bidang reseptif (*receptive field*) yang menyatakan ukuran dari *filter* yang digunakan untuk setiap neuron. *Filter* yang terbentuk akan menelusuri seluruh bidang reseptif pada gambar sehingga akan terjadi *weight sharing* dan menghasilkan *output* yang disebut *feature map*.

Pooling layer adalah lapisan yang memiliki fungsi untuk meminimalkan ukuran fitur dengan memperkecil ukuran data keluaran *convolution layer*, *overfitting*, dan jaringan yang kompleks berkurang. *Pooling layer* terletak setelah *convolutional layer*. *Average pooling* adalah nilai yang didapat dari nilai rata-rata dan *max pooling* adalah nilai yang didapat dari nilai maksimal.

Activation function merupakan sebuah node yang ditambahkan pada akhir output dari setiap *neural network* untuk menentukan keluarannya. *Activation function* terletak sebelum *pooling layer* dan setelah *convolution layer*. Ada beberapa macam *activation function* yang sering digunakan dalam penelitian antara lain *Rectified Linear Unit* (ReLU), *Leaky* ReLU, dan *Parametric ReLU*.

Reactified Linear Unit (ReLU) merupakan *activation function* yang umum digunakan. ReLU ada lapisan aktivasi di CNN yang digunakan untuk meningkatkan tahap pelatihan pada jaringan saraf yang dapat meminimalkan kesalahan. Penggunaan aktivasi ReLU menghasilkan *output* baru dari hasil konvolusi. Ketika nilai *input* memiliki nilai positif, aktivasi ReLU menetapkan nilai *output* akan sama dengan *input*. Sedangkan Ketika nilai *input* negatif maka nilai *output* berubah menjadi 0.

F. Parameter Penelitian

1. *Bit Error Rate* (BER)

Bit Error Rate merupakan perbandingan tingkat *error* antara *watermark* yang sudah diserang setelah ekstraksi dan *watermark* originalnya. BER dapat digunakan untuk menentukan tingkat *ribustness* atau ketahanan dari suatu *watermark* terhadap serangan. Nilai BER berkisar dari 0 sampai 1, dimana semakin kecil nilai BER maka akan semakin *robust* suatu *watermark* terhadap serangan. Nilai BER dapat ditentukan dengan rumus:

$$BER = \frac{Jumlah\ bit\ kesalahan}{Jumlah\ bit\ yang\ disisipkan}\ x\ 100\% \quad (3)$$

2. *Objective Difference Grade* (ODG)

Objective Difference Grade merupakan parameter pengukuran objektif untuk mengukur perbedaan antara sinyal audio asli dengan sinyal audio yang telah disisipkan *watermark yang* memiliki nilai berkisar antara -4 sampai 0 [8]. Nilai PEAQ telah ditetapkan dalam (ITU-R) BS.1387-1 digunakan untuk mengevaluasi nilai ODG yang dihasilkan. Penilaian ODG dapat dilihat pada tabel berikut:

TABEL 2. 1
Skor ODG dan Penjelasannya

ODG	Penjelasan	Kualitas
0.0	Tak Terlihat	Bagus sekali
-1.0	Terlihat, tetapi tidak mengganggu	Bagus
-2.0	Sedikit mengganggu	Seimbang
-3.0	Mengganggu	Sedikit buruk
-4.0	Sangat mengganggu	Buruk

Sumber: Standar (ITU-R) BS. 1387-1

3. *Signal to Noise Ratio* (SNR)

Signal to Noise Ratio merupakan perbandingan antara tingkat *noise* sinyal audio yang ter*watermark* dengan sinyal audio asli. Semakin tinggi nilai SNR maka akan semakin tinggi pula tingkat *imperceptibility*-nya. Perhitungan SNR dilakukan dengan persamaan berikut:

$$SNR = 10\ log_{10} \frac{\sum_{i=0}^{N-1} x^2(n)}{\sum_{i=0}^{N-1} [x_w(n) - x(n)]^2} \quad (4)$$

dimana:
$x_w(n)$ = audio yang diberi *watermark*
$x(n)$ = *host audio*
N = panjang audio (dalam desibel, dB) sebagai nilai SNR

4. *Capacity*

Capacity merepresentasikan jumlah bit yang dapat disisipkan ke dalam *host* audio. Semakin tinggi *capacity* maka jumlah bit yang dapat disisipkan ke dalam *host* audio semakin tinggi. *Capacity* dihitung menggunakan rumus berikut.

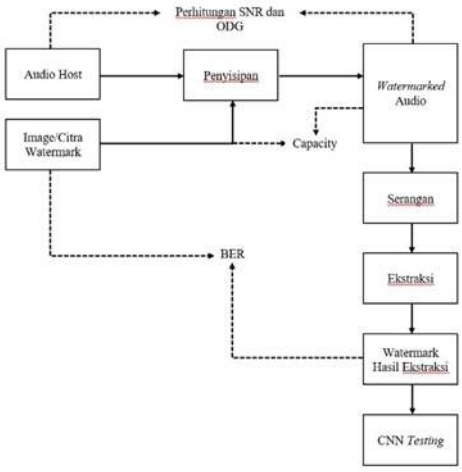
$$C = \frac{Panjang\ bit\ Watermark}{Panjang\ bit\ host}\ x\ f_s \quad (5)$$

dimana:
C = *capacity* dengan satuan *bit per second* (bps)
fs = frekuensi *sampling* dengan satuan sampel per detik (dalam Hertz, Hz)

III. METODE PENELITIAN

Pada penelitian ini, metode yang akan digunakan untuk proses *embedding* adalah Multi-bit SS. Proses *embedding* ini akan menghasilkan audio yang diberi *watermark*. Dalam menentukan kualitas audio yang diberi *watermark* dengan menghitung SNR, ODG, dan *Capacity*.

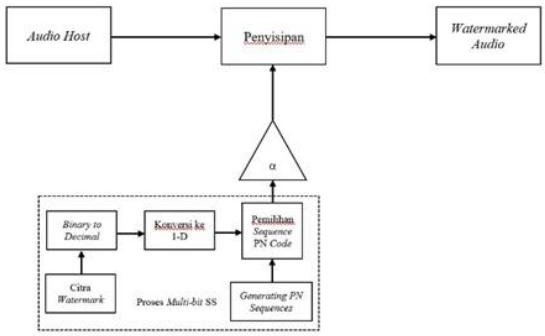
Untuk menentukan ketahanan suatu sistem perlu diberikan serangan. *Audio watermark* yang diserang akan diproses terlebih dahulu sebelum proses ekstraksi menggunakan *Convolutional Neural Network*. Proses ekstraksi merupakan suatu proses untuk memisahkan antara *watermark* dan *audio host*. *Watermark* hasil dari ekstraksi akan dibandingkan dengan *watermark* asli untuk mendapatkan perhitungan BER. Secara umum, blok diagram dari sistem *audio watermarking* dapat dilihat pada Gambar 3.1 berikut ini



GAMBAR 3. 1
Proses Sistem *Audio Watermarking*

A. Proses Penyisipan

Proses penyisipan merupakan proses penyisipan *watermark* ke dalam *audio host*. Adapun langkah-langkah proses penyisipan, sebagai berikut:



GAMBAR 3. 2
Proses Penyisipan *Watermark*

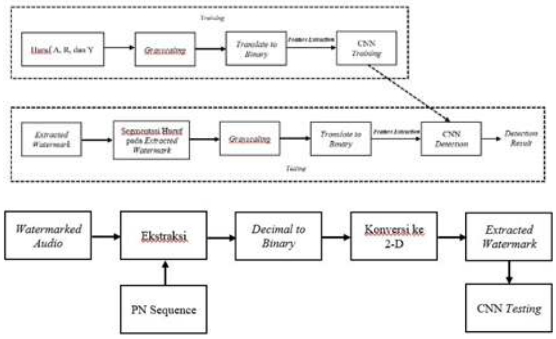
Tahapan dalam proses penyisipan:

1. *Audio host* yang digunakan adalah audio mono agar mempermudah proses penyisipan karena hanya memiliki 1 kanal.
2. Mengkoversi citra *watermark* yang berupa biner menjadi desimal.
3. Citra *watermark* yang sudah dikonversi menjadi desimal kemudian dikonversi lagi menjadi 1 dimensi agar sesuai dengan *audio host*.
4. Pada proses penyisipan perlu melakukan proses *generating PN sequence* dengan beberapa tahapan sebagai berikut:
   a. Menentukan banyaknya PN *Sequence* dengan rumus:
   $$N_p = 2^{N_b} \quad (6)$$
   dimana N_p adalah banyaknya PN *Sequence,* dan N_b adalah jumlah bit *watermark* yang disisipkan.
   b. Lalu asumsikan P_1 sebagai PN *Sequence* dengan N sebagai panjangnya.
   $$P_1 = [p_{11}, p_{12}, p_{13}, \dots, p_{1N}] \quad (7)$$
   dimana $N > N_p$ dan $p_i \in \{-1, +1\}$ dan $i = 1,2,3, \dots, N$.
   c. P_2 hingga P_{NP} dapat ditentukan berdasarkan P_1 yang sudah ditentukan sebelumnya, hasil PN *Sequence* menjadi:
   $$\begin{cases} P_2 = [p_{21}, p_{22}, p_{23}, \dots, p_{2N}] \\ P_3 = [p_{31}, p_{32}, p_{33}, \dots, p_{3N}] \\ \quad \vdots \\ P_{Np} = [p_{Np1}, p_{Np2}, p_{Np3}, \dots, p_{NpN}] \end{cases} \quad (8)$$
5. Membuat representasi antara PN *Sequence* yang telat dibuat dengan data citra *watermark* yang telah dikoversi menjadi 1-D.
6. PN *Sequence* yang diperoleh dari tahap 5 dikalikan dengan α yang nilainya disesuaikan agar mendapatkan hasil yang optimal dari sistem.
7. *Watermark* siap untuk disisipkan ke dalam *audio host* menggunakan persamaan (1) dengan menambahkan sinyal *audio host* yang sudah ditransformasi dengan hasil dari *Multi-bit* SS.

B. Proses Deteksi Menggunakan CNN

Proses deteksi merupakan proses untuk mengetahui *watermark* terkena serangan atau tidak. Secara umum, proses deteksi menggunakan CNN dibagi menjadi 4 tahapan yaitu pengambilan data, *pre-processing*, *training*, dan tahap klasifikasi.

Jumlah data yang akan di*training* adalah 3. Data yang digunakan sebagai input berupa huruf A, R, dan Y yang digunakan sebagai *watermark*. Melatih (*Training*) data untuk mengenali objek dan mengklasifikasi data, kemudian menguji (*Testing*) data untuk mengevaluasi performa model dan menilai hasil prediksi. Proses *training* dan proses *testing* diilustrasikan pada Gambar 3. 4.



GAMBAR 3. 3
Proses Penyisipan Watermark

IV. HASIL DAN PEMBAHASAN

Bab ini membahas analisis dan hasil pengujian *audio watermarking* untuk mengetahui ketahanan dan kualitasnya terhadap serangan. Data *watermark* berupa teks 'ARY'. Teks *watermark* dikonversi menjadi gambar yang menghasilkan *watermark* gambar seperti yang ditunjukkan pada Gambar 4.1. Simulasi *audio watermarking* dirancang menggunakan *host* audio dengan format *.wav, dan frekuensi sampling 44100 Hz. *Host* audio yang digunakan dalam pengujian *audio watermarking* dapat dilihat pada Tabel 4.1.

GAMBAR 4. 1
Watermark

TABEL 4. 1
Host audio

Index	*Host* audio
1	temple_of_love-sisters_of_mercy.wav
2	evangeline-matthew_sweet.wav
3	i_ran_so_far_away-flock_of_seagulls.wav

Untuk menentukan kualitas dari *audio watermarking* yang sudah dirancang, berikut adalah langkah-langkah pengujiannya:

1. Menganalisis hasil simulasi *audio watermarking* tanpa serangan untuk mengetahui pengaruh dari jumlah bit per segmen (jbsf), panjang segmen (LN), dan parameter alfa (α) terhadap *Bit Error Rate* (BER), O*bjective Difference Grade* (ODG), *Signal to Noise Ratio* (SNR), dan *Capacity* (C). Parameter sampel uji memiliki batasan sebagai berikut:
   a. Jbsf: 2,4,6, dan 8
   b. LN: 512, 1024, 2048, dan 4096
   c. α: 0.001 hingga 0.009
2. Menganalisis hasil simulasi *audio watermarking* terhadap serangan dengan optimasi parameter. Tiga *host* audio disimulasi untuk mendapatkan parameter yang optimal. Parameter terbaik dipilih dari nilai BER yang paling mendekati 0, ODG yang lebih besar dari -2, SNR yang paling tinggi, dan C yang paling tinggi dengan akurasi teks 100%.

A. Analisis Pengaruh dari Jumlah Bit per Segmen (jbsf)

Nilai jbsf yang di analisis adalah 2, 4, 6, dan 8. Pengaruh nilai jbsf terhadap BER, ODG, SNR, dan C dapat dilihat pada Tabel 4.2.

TABEL 4. 2
Analisis pengaruh dari jumlah bit per segmen (jbsf)

α	jbsf	LN	ODG	SNR	C	BER	Akurasi Teks
0.005	2	2048	-0.79	18.41	14.36	0.0065	100%
0.005	4	2048	-0.92	17.67	28.71	0.0078	100%

0.005	6	2048	-0.93	17.55	43.07	0.0156	66.67%
0.005	8	2048	-0.98	17.28	57.42	0.0104	100%

Berdasarkan hasil pengujian pada Tabel 4.2, nilai jbsf berpengaruh terhadap BER, ODG, SNR, dan C. Semakin besar nilai jbsf maka semakin besar nilai C. Hal ini disebabkan karena ketika nilai jbsf semakin besar, maka jumlah bit setiap segmen lebih besar dan menghasilkan kapasitas yang lebih besar. Nilai jbsf juga mempengaruhi nilai ODG, SNR, dan BER. Nilai jbsf sebesar 8 digunakan sebagai parameter untuk pengujian selanjutnya karena memiliki nilai C yang paling tinggi dengan akurasi teks 100%.

B. Analisis Pengaruh dari Panjang Segmen (LN)

Nilai LN yang di analisis adalah 512, 1024, 2048, dan 4096. Pengaruh nilai LN terhadap BER, ODG, SNR, dan C dapat dilihat pada Tabel 4.3.

Tabel 4. 3
Analisis pengaruh dari panjang segmen (LN)

α	jbsf	LN	ODG	SNR	C	BER	Akurasi Teks
0.005	8	2048	-1.27	15.37	229.69	0.0872	66.67%
0.005	8	2048	-1.38	16.55	114.84	0.0495	66.67%
0.005	8	2048	-0.98	17.28	57.42	0.0104	100%
0.005	8	2048	-1.08	17.9	28.71	0	100%

Berdasarkan hasil pengujian pada Tabel 4.3, nilai LN berpengaruh terhadap BER, ODG, SNR, dan C. Semakin besar nilai LN maka semakin kecil nilai C. Hal ini disebabkan karena ketika nilai LN semakin besar, maka jumlah sampel lebih kecil yang mengakibatkan bit *watermark* yang tertanam lebih sedikit sehingga nilai C semakin kecil. Nilai LN juga mempengaruhi nilai ODG, SNR, dan BER. Nilai LN sebesar 2048 digunakan sebagai parameter untuk pengujian selanjutnya karena memiliki nilai ODG yang paling baik dengan akurasi teks 100%.

C. Analisis Pengaruh dari Parameter Alfa (α)

Nilai alfa yang di analisis adalah 0.001 sampai 0.009. Pengaruh nilai alfa terhadap BER, ODG, SNR, dan C dapat dilihat pada Tabel 4.4.

TABEL 4. 4
Analisis pengaruh dari parameter alfa (α)

α	jbsf	LN	ODG	SNR	C	BER	Akurasi Teks
0.001	8	2048	-0.08	31.26	57.42	0.4622	0%
0.002	8	2048	-0.35	25.24	57.42	0.2852	0%
0.003	8	2048	-0.61	21.72	57.42	0.1094	33.33%
0.004	8	2048	-0.84	19.22	57.42	0.0313	100%
0.005	8	2048	-0.98	17.28	57.42	0.0104	100%
0.006	8	2048	-1.14	15.7	57.42	0.0046	100%
0.007	8	2048	-1.32	14.36	57.42	0	100%
0.008	8	2048	-1.49	13.2	57.42	0	100%
0.009	8	2048	-1.65	12.18	57.42	0	100%

Berdasarkan hasil pengujian pada Tabel 4.4, nilai alfa berpengaruh terdahap BER, ODG, dan SNR. Semakin besar nilai alfa maka semakin kecil nilai BER, ODG, dan SNR. Semakin besar nilai alfa maka kualitas dari audio yang telah di *watermark* akan semakin tidak baik. Hal ini disebabkan karena ketika menggunakan nilai alfa yang besar, maka bit sinkronisasi akan semakin kuat. Sehingga dapat mempengaruhi dari kualitas audio itu sendiri. Tetapi semakin besar nilai alfa maka semakin kecil nilai BER yang berarti membuat *watermark* menjadi lebih kuat. Nilai alfa sebesar 0.004 adalah yang paling optimal karena memiliki nilai ODG yang paling baik dengan akurasi teks 100%.

D. Parameter Optimal

Setelah menganalisis pengaruh dari masing-masing parameter terhadap kinerja dari *audio watermarking,* parameter yang dipilih adalah parameter yang paling baik untuk pengujian *audio watermarking* terhadap serangan. Parameter optimal dapat dilihat pada Tabel 4.5. Nilai jbsf = 8 dan LN = 2048 akan digunakan untuk analisis ketahanan *audio watermarking* terhadap serangan.

TABEL 4. 5
Optimasi parameter tanpa serangan

α	jbsf	LN	ODG	SNR	C	BER	Akurasi Teks
0.004	8	2048	-0.84	19.22	57.42	0.0313	100%

Parameter optimal dari pengujian *audio watermarking* terhadap serangan kompresi MP3 64 kbps untuk masing-masing *host* audio dapat dilihat pada Tabel 4.6.

TABEL 4. 6
Parameter optimal terhadap serangan

Host audio	α	jbsf	LN
temple_of_love-sisters_of_mercy.wav	0.005	8	2048
evangeline-matthew_sweet.wav	0.004	8	2048
i_ran_so_far_away-flock_of_seagulls.wav	0.006	8	2048

Nilai BER, ODG, SNR, C, dan akurasi teks dari pengujian *audio watermarking* terhadap serangan kompresi MP3 64 kbps untuk masing-masing *host* audio dengan nilai parameter optimal masing-masing dapat dilihat pada Tabel 4.7.

TABEL 4. 7
Hasil pengujian dari parameter optimal terhadap serangan

Host audio	ODG	SNR	C	BER	Akurasi Teks
temple_of_love-sisters_of_mercy.wav	-0.98	17.29	57.42	0.0983	33.33%
evangeline-matthew_sweet.wav	-2.16	13.06	57.42	0.0736	66.67%
i_ran_so_far_away-flock_of_seagulls.wav	-1.41	17.74	57.42	0.1947	33.33%

Berdasarkan Tabel 4.6, menunjukkan bahwa parameter yang optimal adalah nilai jbsf sebesar 8, nilai LN sebesar 2048, dan nilai alfa pada rentang 0.004 sampai 0.006. Hasil pengujian ditunjukkan pada Tabel 4.7 dengan nilai ODG pada rentang -0.9821 hingga -2.1589, nilai SNR pada rentang 13.0614 hingga 17.7351, nilai C sebesar 57.4219, nilai BER pada rentang 0.0736 hingga 0.1947, dan akurasi teks pada rentang 33.33% hingga 66.67%.

E. Analisis Ketahanan *Audio Watermarking* dengan Parameter Optimal Terhadap Serangan

Pada bagian ini akan dilakukan pengujian terhadap ketahanan skema *audio watermarking* yang sudah dirancang terhadap beberapa jenis serangan dengan menggunakan parameter optimal berdasarkan pengujian sebelumnya. Hasil rata-rata BER dan rata-rata akurasi teks dapat dilihat pada Tabel 4.8.

TABEL 4. 8
Rata-rata BER dan akurasi teks terhadap serangan

Host audio	Rata-rata BER	Rata-rata Akurasi Teks
temple_of_love-sisters_of_mercy.wav	0.2533	36.2744%
evangeline-matthew_sweet.wav	0.2238	32.3541%
i_ran_so_far_away-flock_of_seagulls.wav	0.3006	23.5294%

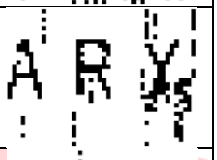
Berdasarkan Tabel 4.8, nilai rata-rata BER adalah 0.2533 dan rata-rata akurasi teks adalah 36.2744% untuk *host* audio temple_of_love-sisters_of_mercy.wav, nilai rata-rata BER adalah 0.2238 dan rata-rata akurasi teks adalah 32.3541% untuk *host* audio evangeline-matthew_sweet.wav, nilai rata-rata BER adalah 0.3006 dan rata-rata akurasi teks adalah 23.5294% untuk *host* audio i_ran_so_far_away-flock_of_seagulls.wav.

F. *Extracted Watermark*

Proses ekstraksi *watermark* dari *watermarked audio* menghasilkan nilai BER dan akurasi teks yang berbeda untuk setiap *host* audio. Hasil ekstraksi *watermark* dapat dilihat pada

Tabel 4.9. Tabel 4.9 menunjukkan bahwa nilai BER sebesar 0 dan akurasi teks 100% menghasilkan *extracted watermark* yang sempurna. Semakin tinggi nilai BER maka semakin buruk hasil *extracted watermark*-nya. Semakin buruk hasil *extracted watermark* mengakibatkan CNN gagal untuk mendeteksi setiap huruf maka akurasi teks semakin kecil.

TABEL 4. 9
Extracted Watermark

BER	Akurasi Teks	Serangan	*Extracted Watermak*
0.263	0%	Low Pass Filter 6000 Hz	
0.153	33.33%	Band Pass Filter 100-9000 Hz	
0.0684	66.67%	Echo	
0.0254	100%	Kompresi MP3 128 kbps	
0	100%	Tanpa Serangan	

G. Perbandingan dengan Penelitian Terkait

Pada subbab ini akan membahas perbandingan dengan penelitian terkait yang sudah dilakukan sebelumnya. Tabel 4.10 menunjukkan perbandingan ODG, SNR, C, dan BER dari skema yang diusulkan dengan penelitian terkait [1], [3], [10].

TABEL 4. 10
Perbandingan dengan Penelitian Terkait

Parameter		Referensi			
		[1]	[3]	[10]	Skema yang diusulkan
ODG		N/A	-0.22	-2.74	-1.52
SNR		N/A	22.64	21.87	16.03
C		689.06	83.00	6.67	57.42
BER	LPF (9k)	0.00	N/A	0.03	0.24
	Noise (0 dB)	0.43	N/A	0.13	0.05
	Noise (20 dB)	0.126	0	0.03	0.03
	Resampling (22 kHz)	0.04	0	0.03	0.20
	Linear Speed Change (5%)	0.00	N/A	0.05	0.05
	MP3 (64 kbps)	0.16	0.00	0.03	0.12

Skema yang diusulkan adalah skema yang digunakan pada penelitian ini dan N/A merupakan tidak tersedia atau tidak ada data pada jurnal referensi. Skema yang diusulkan memiliki nilai ODG yang lebih tinggi dibandingkan dengan [10], tetapi lebih rendah dibandingkan dengan [3]. Nilai SNR dari skema yang diusulkan memiliki nilai rata-rata sebesar 16.03 yang lebih rendah dibandingkan dengan [3] dan [10]. Sementara itu, *capacity* dari skema yang diusulkan memiliki nilai yang lebih tinggi dibandingkan dengan [10], tetapi lebih rendah dibandingkan dengan [1] dan [3].

Perbandingan nilai rata-rata BER antara skema yang diusulkan dengan metode-metode sebelumnya terhadap beberapa serangan ditunjukkan juga pada Tabel 4.10. Metode yang diusulkan menghasilkan nilai BER yang lebih tinggi dibandingkan dengan metode sebelumnya terhadap serangan LPF (9k) dan *resampling* (22 kHz). Tetapi metode yang diusulkan menghasilkan nilai BER yang lebih rendah dibandingkan dengan [1] dan [10] terhadap serangan Noise (0 dB) dengan nilai BER sebesar 0.05.

V. KESIMPULAN

A. Kesimpulan

Setelah melakukan pengujian *audio watermarking* dengan metode *Multi-bit SS* dan *Convolutional Neural Network* (CNN) terhadap beberapa serangan, maka dapat disimpulkan sebagai berikut:

1. Skema *audio watermarking* mendapatkan hasil terbaik dengan menggunakan parameter optimal yaitu jbsf sebesar 8, LN sebesar 2048 dan nilai alfa (α) dari 0.004 sampai 0.006.
2. Skema *audio watermarking* menggunakan metode *Multi-bit SS* dan *Convolutional Neural Network* (CNN) diuji tanpa menggunakan serangan menghasilkan nilai rata-rata *Signal to Noise Ratio* (SNR) sebesar 16.03, nilai rata-rata *Objective Difference Grade* (ODG) sebesar -1.52, nilai *capacity* (C) sebesar 57.42, nilai rata-rata *Bit Error Rate* (BER) sebesar 0.016, dan nilai akurasi teks sebesar 100%.
3. Ketahanan skema *audio watermarking* diuji dengan menggunakan serangan. Nilai rata-rata *Bit Error Rate* (BER) terhadap semua serangan sebesar 0.2592 dengan nilai terendah sebesar 0.2533 dan nilai tertinggi sebesar 0.3006. Nilai rata-rata akurasi teks terhadap semua serangan sebesar 30.72% dengan nilai terendah sebesar 23.53% dan nilai tertinggi sebesar 36.28%.
4. Skema *audio watermarking* yang diusulkan dibandingkan dengan metode-metode sebelumnya. Hasilnya menunjukkan bahwa skema yang diusulkan menghasilkan nilai BER yang lebih rendah dibandingkan dengan metode sebelumnya terhadap serangan Noise (0 dB). Tetapi menghasilkan nilai BER yang lebih tinggi dibandingkan dengan metode sebelumnya terhadap serangan LPF (9k) dan *resampling* (22 kHz).

B. Saran

Dari analisis yang telah dijelaskan terdapat beberapa kendala yang mungkin bisa dikembangkan atau diperbaiki untuk menghasilkan skema *audio watermarking* lebih baik lagi. Oleh karena itu, saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Disarankan untuk penelitian selanjutnya menggunakan skema *blind audio watermarking*.
2. Disarankan untuk menggunakan arsitektur *machine learning* yang berbeda.
3. Disarankan untuk menambahkan metode lain pada proses penyisipan dan ekstraksi *watermark*.

REFERENSI

[1] G. BUDIMAN, S. AULIA, and I. N. A. RAMATRYANA, "Penyisipan Citra pada Audio dengan Kode PN Terdistribusi Gaussian," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 7, no. 2, p. 209, May 2019, doi: 10.26760/elkomika.v7i2.209.

[2] Y. Xiang, I. Natgunanathan, D. Peng, G. Hua, and B. Liu, "Spread Spectrum Audio Watermarking Using Multiple Orthogonal PN Sequences and Variable Embedding Strengths and Polarities," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 26, no. 3, pp. 529–539, Mar. 2018, doi: 10.1109/TASLP.2017.2782487.

[3] A. A. Attari, A. Asghar, and B. Shirazi, "Robust and Transparent Audio Watermarking based on Spread Spectrum in Wavelet Domain."

[4] G. Budiman, A. B. Suksmono, and D. Danudirdjo, "Compressive sampling with multiple bit spread spectrum-based data hiding," *Applied Sciences (Switzerland)*, vol. 10, no. 12, Jun. 2020, doi: 10.3390/app10124338.

[5] A. Alzahrani, "Detecting Digital Watermarking Image Attacks Using a Convolution Neural Network Approach," *Security and Communication Networks*, vol. 2022, 2022, doi: 10.1155/2022/4408336.
[6] A. Das and X. Zhong, "A Deep Learning-based Audio-in-Image Watermarking Scheme," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.02436
[7] J. E. Lee, Y. H. Seo, and D. W. Kim, "Convolutional neural network-based digital image watermarking adaptive to the resolution of image and watermark," *Applied Sciences (Switzerland)*, vol. 10, no. 19, Oct. 2020, doi: 10.3390/app10196854.
[8] Surya Engineering College and Institute of Electrical and Electronics Engineers, *Proceedings of the 3rd International Conference on Computing Methodologies and Communication (ICCMC 2019) : 27-29, March 2019*.
[9] Institute of Electrical and Electronics Engineers., *[4th International Conference on Computer and Communication Engineering] : [Kuala Lumpur, July 3-5, 2012, ICCCE 2012]*. IEEE, 2012.
[10] R. Naufal Alief, G. Budiman, and L. Novamizanti, *Audio Watermarking Berbasiskan DWT-DCT Menggunakan Multibit Spread Spectrum Audio Watermarking Based on DWT-DCT Using Multibit Spread Spectrum*. 2019.
[11] A. Tavakoli, Z. Honjani, and H. Sajedi, "Convolutional Neural Network-Based Image Watermarking using Discrete Wavelet Transform," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.06179