

**KLASIFIKASI INSTRUMEN MELODIS PADA KOREAN
BALLAD INSTRUMENTAL DENGAN CNN DAN
SPECTROGRAM**

TUGAS AKHIR

Diajukan sebagai syarat menyelesaikan jenjang strata Satu (S-1) di
Program Studi Teknik Informatika, Fakultas Teknologi Industri, Institut
Teknologi Sumatera

Oleh:

Ramon Riping

122140078



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2026**

DAFTAR ISI

DAFTAR ISI	ii
DAFTAR TABEL	iii
DAFTAR GAMBAR	iv
DAFTAR RUMUS	v
DAFTAR KODE	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Tinjauan Pustaka	7
2.2 Dasar Teori	9
2.2.1 Spectrogram	9
2.2.2 Convolutional Neural Network (CNN)	9
2.2.3 Feature Extraction Audio	9
2.2.4 Korean Ballad	9
BAB III METODE PENELITIAN	10
3.1 Alur Penelitian	10
3.2 Penjabaran Langkah Penelitian	11
3.2.1 Pengumpulan Dataset	11

3.2.2	Preprocessing	11
3.2.3	Konversi ke Spectrogram	11
3.2.4	Pelatihan Model CNN	11
3.2.5	Evaluasi	11
3.3	Alat dan Bahan Tugas Akhir	11
3.3.1	Alat	11
3.3.2	Bahan	12
3.4	Metode Pengembangan	12
3.4.1	Alur Pengembangan Tugas Akhir	12
3.4.2	Cara Pengumpulan Data	12
3.4.3	Metode Pengembangan Tugas Akhir	13
3.4.4	Metode Pengujian Penelitian	13
3.5	Ilustrasi Perhitungan Metode	14
3.5.1	Representasi Spectrogram	14
3.5.2	Prediksi Model CNN	14
3.5.3	Perhitungan Akurasi	15
3.6	Rancangan Pengujian	15
DAFTAR PUSTAKA		16

DAFTAR TABEL

Tabel 2.1	Literasi Penelitian Terdahulu	7
Tabel 3.1	Contoh Hasil Prediksi Model CNN	14

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	10
Gambar 3.2 Alur Pengembangan Tugas Akhir	12

DAFTAR RUMUS

Rumus 3.1	Rumus Accuracy	15
Rumus 3.2	Contoh Perhitungan Accuracy	15

DAFTAR KODE

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam konteks mobilitas fisik di daerah perkotaan, indra pendengaran merupakan mekanisme deteksi alami yang cukup penting untuk mengetahui kondisi lingkungan di sekitar. Namun, fungsi indra tersebut tidak dimiliki oleh penyandang Tuna Rungu, yang hanya bisa mengandalkan penglihatan mereka untuk memantau keadaan. Ketergantungan penuh pada aspek visual ini dapat menjadi kerentanan serius, mengingat mereka memiliki keterbatasan sudut pandang dan tidak dapat memantau kondisi di luar jangkauan penglihatan. Akibatnya, ancaman yang muncul dari titik buta (*blind spot*), seperti gonggongan anjing yang mengejar atau klakson kendaraan yang melaju kencang dari arah belakang, seringkali terlambat disadari, sehingga berpotensi menimbulkan kecelakaan fatal.

Saat ini, Alat Bantu Dengar (ABD) merupakan perangkat yang umum digunakan untuk menunjang komunikasi verbal penyandang Tuna Rungu. Meskipun efektif untuk komunikasi verbal jarak dekat, alat ini memiliki keterbatasan signifikan dalam konteks keselamatan di luar ruangan. Hal ini disebabkan oleh penurunan selektivitas frekuensi (*reduced frequency selectivity*) yang umum terjadi pada gangguan pendengaran sensorineural, sehingga menyulitkan pemisahan sinyal suara utama dari kebisingan latar belakang yang tumpang tindih [1]. Kondisi ini diperburuk oleh keterbatasan teknis ABD, di mana sekadar amplifikasi sinyal suara tidak cukup untuk mengembalikan kemampuan pemilahan suara secara alami. Akibatnya, sinyal ancaman penting seringkali tertutup oleh suara-suara lainnya, yang berdampak pada hilangnya kewaspadaan situasional pengguna.

Untuk mengatasi kendala tersebut, teknologi klasifikasi suara lingkungan

(*Environmental Sound Classification/ESC*) berbasis *Deep Learning* hadir sebagai solusi bagi para penyandang Tuna Rungu dalam mengenali suara-suara bahaya di sekitar mereka. Dengan memanfaatkan model *Deep Learning* yang telah dilatih untuk mengenali berbagai jenis suara lingkungan, sistem ESC dapat mendeteksi dan mengklasifikasikan suara-suara bahaya seperti sirene, klakson mobil, gonggongan anjing, atau suara tembakan. Setelah suara terdeteksi dan diklasifikasikan, sistem dapat memberikan notifikasi dalam bentuk visual atau getaran kepada pengguna Tuna Rungu, sehingga mereka dapat segera menyadari ancaman dan mengambil tindakan antisipasi. Dengan demikian, teknologi ESC berbasis *Deep Learning* memiliki potensi besar untuk meningkatkan keselamatan dan kualitas hidup penyandang Tuna Rungu dalam mobilitas sehari-hari di lingkungan perkotaan.

Namun, pengembangan model ESC yang mampu mendeteksi bahaya secara akurat menghadapi tantangan teknis, terutama terkait kebutuhan dataset berskala besar. Oleh karena itu, metode *Transfer Learning* menjadi solusi efektif untuk melatih model deteksi, dengan memanfaatkan kerangka kerja *Pre-trained Audio Neural Networks* (PANNs). PANNs dipilih karena tidak hanya menyediakan bobot model dari dataset masif [2], tetapi juga menyediakan variasi arsitektur yang dirancang khusus untuk menangani berbagai representasi input audio. Hal ini memunculkan urgensi untuk mengevaluasi arsitektur mana yang paling optimal, apakah arsitektur yang bekerja pada domain waktu dengan input *Raw Waveform* (satu dimensi), atau arsitektur pada domain frekuensi dengan input *Spectrogram* (dua dimensi). Selain itu, PANNs juga menyediakan arsitektur dengan pendekatan *Hybrid* yang menggabungkan kedua representasi tersebut, yang secara teoritis berpotensi memaksimalkan akurasi deteksi.

Meskipun Kong et al. [2] telah memaparkan tolok ukur kinerja model-model tersebut pada dataset masif AudioSet, performa tersebut belum tentu linier ketika diterapkan pada *downstream task* dengan ketersediaan data yang terbatas (*data scarcity*) seperti pada kasus klasifikasi suara lingkungan perkotaan.

Kompleksitas arsitektur model *Hybrid* dan *Spectrogram*, misalnya, memiliki risiko *overfitting* yang lebih tinggi dibandingkan model *Raw Waveform* ketika dilatih pada dataset yang kecil. Oleh karena itu, penelitian ini menjadi krusial untuk mengevaluasi ulang adaptabilitas dan generalisasi ketiga arsitektur tersebut secara spesifik pada dataset UrbanSound8K.

Guna memastikan tingkat akurasi deteksi yang maksimal bagi keselamatan pengguna, penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja dari tiga pendekatan representasi input, yaitu: *Raw Waveform*, *Spectrogram*, dan *Hybrid*. Melalui perbandingan ini, diharapkan dapat ditemukan model terbaik yang tidak hanya memiliki akurasi tinggi, tetapi juga mampu meminimalisir kesalahan deteksi yang fatal. Hasil akhir dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan teknologi asistif berbasis AI yang andal, guna meningkatkan kewaspadaan dan jaminan keselamatan bagi komunitas Tuna Rungu.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pengaruh perbedaan representasi input (*Raw Waveform*, *Spectrogram*, dan *Hybrid*) terhadap performa model *Pre-trained Audio Neural Networks* (PANNs) dalam mengklasifikasikan suara bahaya pada kondisi ketersediaan data yang terbatas?
2. Representasi input manakah yang menghasilkan model paling optimal (berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*) untuk meminimalisir kesalahan deteksi pada sistem keselamatan penyandang Tuna Rungu?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menganalisis pengaruh perbedaan representasi input (*Raw Waveform*, *Spectrogram*, dan *Hybrid*) terhadap performa model *Pre-trained Audio Neural Networks* (PANNs) dalam mengklasifikasikan suara bahaya pada kondisi ketersediaan data yang terbatas.
2. Menentukan pendekatan representasi input yang menghasilkan model paling optimal (berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*) untuk meminimalisir kesalahan deteksi pada sistem keselamatan penyandang Tuna Rungu.

1.4 Batasan Masalah

Batasan masalah yang didefinisikan dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan dataset publik *UrbanSound8K* sebagai representasi suara lingkungan perkotaan. Penulis tidak melakukan pengambilan data primer (perekaman langsung di lapangan).
2. Fokus klasifikasi dibatasi pada kelas suara yang berpotensi menjadi ancaman atau peringatan bagi keselamatan fisik di jalan raya, yang diambil dari taksonomi dataset *UrbanSound8K* (*car_horn*, *siren*, *dog_bark*, *gun_shot*, dan *drilling*).
3. Penelitian ini berfokus pada eksperimen pelatihan (*training*) dan evaluasi performa model *Deep Learning* secara *offline*. Penelitian ini **tidak mencakup** pembuatan perangkat keras (*hardware*), pengembangan aplikasi antarmuka pengguna (*User Interface*) berbasis *mobile/web*, maupun implementasi sistem secara *real-time*.
4. Kerangka kerja yang digunakan adalah *Pre-trained Audio Neural Networks* (PANNs) dengan arsitektur spesifik: *ResIdNet31* (untuk *Raw Waveform*), *CNN14* (untuk *Spectrogram*), dan *Wavegram_Logmel-CNN14* (untuk *Hybrid*).
5. Kinerja model diukur berdasarkan parameter kuantitatif meliputi *Accuracy*,

Precision, Recall, dan F1-Score.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Memberikan bukti empiris terkait efektivitas metode *Transfer Learning* pada arsitektur PANNs serta perbandingan performa antara representasi input *Raw Waveform*, *Spectrogram*, dan *Hybrid* dalam mengatasi keterbatasan dataset (*limited dataset*).
2. Berkontribusi dalam pengembangan teknologi asistif berbasis AI yang dapat meningkatkan keselamatan dan kemandirian mobilitas penyandang Tuna Rungu melalui deteksi suara bahaya yang akurat.
3. Menjadi referensi bagi penelitian selanjutnya atau pengembang aplikasi dalam menentukan konfigurasi model yang paling optimal untuk diterapkan pada sistem peringatan dini.

1.6 Sistematika Penulisan

Sistematika penulisan berisi pembahasan apa yang akan ditulis disetiap Bab. Sistematika pada umumnya berupa paragraf yang setiap paragraf mencerminkan bahasan setiap Bab.

Bab I

Bab ini berisikan penjelasan latar belakang dari topik penelitian yang berlangsung, rumusan masalah dari masalah yang dihadapi pada penjelasan di latar belakang, tujuan dari penelitian, batasan dari penelitian, manfaat dari hasil penelitian, dan sistematika penulisan tugas akhir.

Bab II

Bab ini membahas mengenai tinjauan pustaka dari penelitian terdahulu dan dasar teori yang berkaitan dengan penelitian ini.

Bab III

Bab ini berisikan penjelasan alur kerja sistem, alat dan data yang digunakan, metode yang digunakan, dan rancangan pengujian.

Bab IV

Bab ini membahas hasil implementasi dan pengujian dari penelitian yang dilakukan, serta analisis dan evaluasi yang dapat dipetik dari hasil.

Bab V

Bab ini membahas kesimpulan dari hasil penelitian dan juga saran untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Tinjauan pustaka dijelaskan dalam Tabel 2.1 dan digunakan sebagai referensi berdasarkan penelitian yang berkaitan dengan penelitian ini.

Tabel 2.1 Literasi Penelitian Terdahulu

No.	Judul [Tahun]	Masalah	Metode	Hasil
1.	Deep Spectrogram Learning for Gunshot Classification: A Comparative Study of CNN Architectures and Time-Frequency Representations [2025]	Metode tradisional sering gagal membedakan jenis senjata api berdasarkan suara tembakan secara seakurat karena perbedaan fisik yang halus pada senjata api.	CNN dengan input Mel-Spectrogram; preprocessing menggunakan Librosa, pelatihan model dilakukan dengan Keras	Sistem dapat mengklasifikasikan 6 jenis alat musik secara akurat dengan nilai akurasi tertinggi mencapai 94%

No.	Judul [Tahun]	Masalah	Metode	Hasil
2.	Deteksi Suara Chord Piano Menggunakan Metode Convolutional Neural Network (2022)	Sulitnya mendeteksi akor piano secara otomatis dalam format suara	CNN + Mel-spectrogram	Model CNN mampu mendeteksi akor mayor dan minor piano dengan akurasi 85%
3.	Pengembangan Aplikasi Klasifikasi Suara Alat Musik Kalimba (2024)	Belum ada sistem klasifikasi otomatis suara alat musik kalimba	CNN + Mel-spectrogram	Aplikasi berhasil mengklasifikasi suara kalimba ke beberapa jenis nada dengan performa tinggi
4.	Recognising Bonang Barung Gamelan Instrument Playing Technique Using CNN (2025)	Kurangnya metode otomatis untuk mengenali teknik permainan bonang barung	CNN + Mel-spectrogram	Sistem mampu membedakan teknik pukulan bonang dengan akurasi yang cukup baik

No.	Judul [Tahun]	Masalah	Metode	Hasil
5.	Klasifikasi Suara Instrumen Musik Tiup Menggunakan Metode CNN (2024)	Sulitnya mengidentifikasi jenis instrumen musik tiup dari file audio	CNN + Mel-spectrogram	Model berhasil mengklasifikasi berbagai instrumen tiup (saxophone, flute, trumpet) secara akurat

2.2 Dasar Teori

2.2.1 Spectrogram

Spectrogram merupakan representasi visual dari frekuensi sinyal audio terhadap waktu.

2.2.2 Convolutional Neural Network (CNN)

Convolutional Neural Network atau CNN adalah arsitektur jaringan saraf dalam pembelajaran mendalam, digunakan untuk klasifikasi citra dan spectrogram.

2.2.3 Feature Extraction Audio

Feature Extraction Audio merupakan proses mengambil fitur penting seperti mel-frequency cepstral coefficients (MFCC), chroma, atau spectrogram.

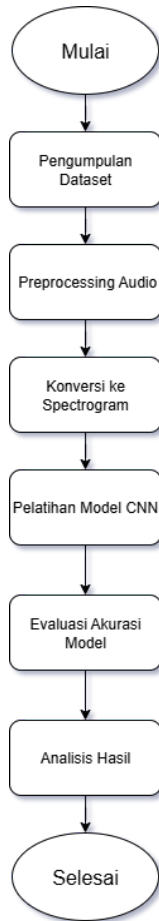
2.2.4 Korean Ballad

Korean Ballad adalah genre musik asal Korea yang dikenal lembut, emosional, dan memiliki aransemen instrumental kuat.

BAB III

METODE PENELITIAN

3.1 Alur Penelitian



Gambar 3.1 Alur Penelitian

3.2 Penjabaran Langkah Penelitian

3.2.1 Pengumpulan Dataset

Mengumpulkan audio instrumental Korean Ballad dari sumber daring (bebas lisensi) atau dataset open-source.

3.2.2 Preprocessing

Melakukan normalisasi audio (mono, sampling rate), segmentasi, dan trimming.

3.2.3 Konversi ke Spectrogram

Mengubah audio menjadi citra spectrogram menggunakan Short-Time Fourier Transform (STFT).

3.2.4 Pelatihan Model CNN

Melatih CNN pada dataset citra spectrogram untuk klasifikasi instrumen.

3.2.5 Evaluasi

Melakukan pengujian terhadap data uji dan mengevaluasi metrik seperti akurasi dan confusion matrix.

3.3 Alat dan Bahan Tugas Akhir

Berisi alat-alat dan bahan-bahan yang digunakan dalam penelitian.

3.3.1 Alat

1. Laptop dengan spesifikasi minimum RAM 16GB, GPU opsional (NVIDIA).
2. Visual Studio Code.
3. Python 3.10 dan library: librosa, matplotlib, numpy, tensorflow.
4. Google Colab
5. Github

3.3.2 Bahan

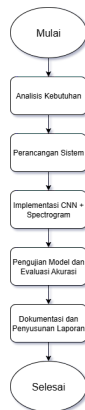
Bahan yang digunakan/diperlukan untuk melakukan penelitian, dapat berupa:

1. Dataset audio instrumen melodis Korean Ballad (open-source atau dikumpulkan sendiri).
2. Dokumentasi teori CNN dan spectrogram.
3. Referensi dari jurnal ilmiah.

3.4 Metode Pengembangan

Penelitian ini menggunakan pendekatan rekayasa perangkat lunak berbasis data untuk mengembangkan sistem klasifikasi instrumen melodis pada musik instrumental Korean Ballad menggunakan metode Convolutional Neural Network (CNN) dan representasi spectrogram. Adapun metode pengembangan dalam penelitian ini dijelaskan sebagai berikut:

3.4.1 Alur Pengembangan Tugas Akhir



Gambar 3.2 Alur Pengembangan Tugas Akhir

3.4.2 Cara Pengumpulan Data

Data dikumpulkan dengan cara sebagai berikut:

1. Dataset diperoleh dari platform seperti YouTube, Freesound, atau sumber dataset open-source.
2. Setiap potongan audio diklasifikasikan secara manual berdasarkan jenis instrumen utama yang terdengar (misalnya: piano, gitar, biola).
3. Untuk meningkatkan akurasi pelabelan, hasil labeling dapat divalidasi oleh seseorang yang memiliki latar belakang musik.

3.4.3 Metode Pengembangan Tugas Akhir

Penelitian ini menggunakan metode pengembangan Waterfall karena setiap tahapan dilakukan secara sistematis dan berurutan, dimulai dari pengumpulan data hingga analisis hasil. Tahapan Waterfall yang digunakan adalah sebagai berikut:

1. Analisis kebutuhan: Menentukan kebutuhan data dan sistem klasifikasi.
2. Desain sistem: Mendesain arsitektur CNN dan proses transformasi audio.
3. Implementasi: Melatih model CNN dengan dataset yang telah disiapkan.
4. Pengujian: Menguji model menggunakan data uji untuk memperoleh metrik evaluasi.
5. Pemeliharaan: Menyesuaikan model atau data jika terdapat ketidaksesuaian hasil.

3.4.4 Metode Pengujian Penelitian

Model CNN akan diuji menggunakan teknik Hold-Out Validation dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Pengujian dilakukan dengan mengukur:

1. Akurasi: Persentase data yang diklasifikasikan dengan benar.
2. Confusion Matrix: Untuk melihat distribusi kesalahan dan keberhasilan klasifikasi antar label.
3. Precision dan Recall: Untuk mengukur performa klasifikasi tiap label secara lebih mendalam.

3.5 Ilustrasi Perhitungan Metode

Pada bagian ini dijelaskan ilustrasi proses klasifikasi instrumen musik melodis dengan menggunakan metode *Convolutional Neural Network* (CNN) dan representasi spectrogram. Ilustrasi dilakukan pada sampel data audio dengan panjang potongan 10 detik.

3.5.1 Representasi Spectrogram

Salah satu file audio Korean Ballad dengan label *biola* diubah ke dalam bentuk spectrogram menggunakan transformasi Short-Time Fourier Transform (STFT). Hasil transformasi berupa matriks dua dimensi dengan sumbu waktu dan frekuensi, yang kemudian digunakan sebagai input untuk model CNN.

3.5.2 Prediksi Model CNN

Spectrogram dari audio dimasukkan ke dalam model CNN. Arsitektur CNN yang digunakan terdiri dari beberapa lapisan konvolusi dan pooling, diakhiri dengan lapisan dense sebagai klasifikasi.

Contoh hasil prediksi dari 10 data uji ditunjukkan pada Tabel 3.1 berikut:

Tabel 3.1 Contoh Hasil Prediksi Model CNN

Data Ke-	Label Sebenarnya	Prediksi Model
1	Biola	Biola
2	Piano	Biola
3	Gitar	Gitar
4	Biola	Biola
5	Piano	Piano
6	Gitar	Gitar
7	Biola	Biola
8	Gitar	Piano
9	Piano	Piano
10	Biola	Biola

3.5.3 Perhitungan Akurasi

Berdasarkan hasil pada Tabel 3.1, diketahui bahwa terdapat 9 prediksi yang benar dari 10 data uji. Akurasi model dihitung menggunakan rumus:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{Rumus 3.1})$$

Misalnya model mengklasifikasikan dengan benar 9 dari 10 data, maka akurasi dihitung sebagai berikut:

$$\text{Accuracy} = \frac{9}{10} = 0.9 = 90\% \quad (\text{Rumus 3.2})$$

Dengan demikian, model CNN dengan input spectrogram pada ilustrasi ini memiliki akurasi sebesar 90% terhadap data uji tersebut.

3.6 Rancangan Pengujian

1. Pengujian Fungsional: apakah model dapat mengenali instrumen piano, flute, dan biola.
2. Pengujian Non-Fungsional: kecepatan pelatihan model dan ukuran file model.
3. Hipotesis: CNN dapat mencapai akurasi minimal 80% dalam klasifikasi 3 instrumen.

DAFTAR PUSTAKA

- [1] Brian C. J. Moore. “Perceptual Consequences of Cochlear Hearing Loss and their Implications for the Design of Hearing Aids”. *Ear and Hearing* 17.2 (1996), pp. 133–161.
- [2] Qiuqiang Kong et al. “PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition”. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), pp. 2880–2894.