

# **JUDUL MENYUSUL**

## **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

<b>DAFTAR ISI</b> .....	<b>ii</b>
<b>DAFTAR TABEL</b> .....	<b>iii</b>
<b>DAFTAR GAMBAR</b> .....	<b>iv</b>
<b>DAFTAR RUMUS</b> .....	<b>v</b>
<b>DAFTAR KODE</b> .....	<b>vi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Batasan Masalah .....	5
1.5 Manfaat Penelitian .....	5
1.6 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
2.1 Tinjauan Pustaka .....	7
2.2 Dasar Teori .....	7
2.2.1 Spectrogram .....	7
2.2.2 Convolutional Neural Network (CNN) .....	7
2.2.3 Feature Extraction Audio .....	7
2.2.4 Korean Ballad .....	7
<b>BAB III METODE PENELITIAN</b> .....	<b>8</b>
3.1 Alur Penelitian .....	8
3.2 Penjabaran Langkah Penelitian .....	8
3.2.1 Identifikasi Permasalahan .....	8

3.2.2	Studi Literatur .....	9
3.2.3	Pengumpulan dan Pra-pemrosesan Data .....	9
3.2.4	Perancangan dan Adaptasi Model .....	10
3.2.5	Eksperimen dan Evaluasi Model .....	11
3.2.6	Analisis Hasil dan Penarikan Kesimpulan .....	12
3.3	Alat dan Bahan Tugas Akhir .....	12
3.3.1	Alat .....	12
3.3.2	Bahan .....	13
3.4	Metode Pengembangan .....	14
3.4.1	Prosedur Eksperimen .....	14
3.4.2	Metode Pengumpulan Data .....	14
3.4.3	Strategi <i>Transfer Learning</i> .....	15
3.4.4	Metode Pengujian Penelitian .....	15
3.5	Ilustrasi Perhitungan Metode .....	16
3.5.1	Ilustrasi Input dan Output Model .....	16
3.5.2	Ilustrasi Perhitungan Evaluasi .....	17
3.6	Rancangan Pengujian dan Hipotesis .....	17
<b>DAFTAR PUSTAKA .....</b>		<b>19</b>

## DAFTAR TABEL

Tabel 3.1	Contoh Data Uji dan Hasil Prediksi.....	17
-----------	---	----

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	8
----------------------------------	---

## DAFTAR RUMUS

## **DAFTAR KODE**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Keselamatan berlalu lintas di lingkungan perkotaan merupakan tantangan krusial yang dihadapi masyarakat saat ini, terutama bagi kelompok rentan seperti penyandang disabilitas [1]. Dalam aktivitas tersebut, indra pendengaran berperan sebagai mekanisme deteksi alami yang 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 akibat tidak adanya peringatan suara. Keterlambatan respon inilah yang secara signifikan meningkatkan risiko terjadinya kecelakaan fatal [2]. Maka dari itu, diperlukan mekanisme bantu yang dapat menggantikan peran indra pendengaran dalam mendeteksi ancaman yang muncul dari luar jangkauan visual.

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 [3]. 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. Keterbatasan perangkat keras dalam memilah sinyal suara ini memunculkan kebutuhan teknologi bagi penyandang Tuna Rungu agar dapat mengidentifikasi suara bahaya melalui pola sinyal suara, dan bukan sekadar amplifikasi sinyal.

Guna mengatasi keterbatasan ini, dikembangkanlah metode cerdas yang dikenal sebagai Klasifikasi Suara Lingkungan atau *Environmental Sound Classification* (ESC). Integrasi teknologi ini pada alat bantu dengar telah lama diteliti sebagai upaya meningkatkan kesadaran situasi pengguna [4]. Pada tahap awal pengembangannya, sistem ESC umumnya dibangun menggunakan metode *Machine Learning* konvensional seperti *Support Vector Machine* (SVM) atau *Random Forest* [5]. Namun, metode-metode klasik tersebut sangat bergantung pada proses ekstraksi fitur secara manual (*hand-crafted features*) yang kaku, sehingga performanya cenderung menurun drastis ketika dihadapkan dengan variasi kebisingan lingkungan yang dinamis. Kelemahan metode tersebut memicu pergeseran tren penelitian menuju pendekatan *Deep Learning*, khususnya *Convolutional Neural Networks* (CNN) yang menawarkan kemampuan untuk mempelajari fitur suara secara otomatis dan hirarkis langsung dari data [6]. Kemampuan adaptasi fitur inilah yang menjadikannya sebagai solusi yang jauh lebih andal dibandingkan metode konvensional. Walaupun menjanjikan akurasi yang lebih tinggi, metode *Deep Learning* membutuhkan dataset berskala masif untuk melatih fitur-fitur tersebut secara efektif. Ketergantungan ini menjadi kendala signifikan pada kasus dengan ketersediaan data yang terbatas, sehingga diperlukan strategi pembelajaran khusus agar model tetap memiliki performa yang *robust*.

Sebagai implementasi strategi tersebut, metode *Transfer Learning* menjadi solusi efektif untuk mengatasi kelangkaan data. Pendekatan ini memanfaatkan

*Pre-trained Audio Neural Networks* (PANNs), yaitu sebuah kerangka kerja model *Deep Learning* skala besar yang telah dilatih sebelumnya (*pre-trained*) pada dataset AudioSet untuk mengenali berbagai pola suara umum [7]. Salah satu keunggulan PANNs terletak pada variasi arsitektur yang dirancang khusus untuk menangani dua jenis representasi input audio yang berbeda. Pertama adalah arsitektur berbasis satu dimensi yang mengolah *Raw Waveform*, yaitu sinyal gelombang suara mentah dalam domain waktu. Kedua adalah arsitektur berbasis dua dimensi yang memanfaatkan *Log-mel Spectrogram*, yaitu representasi visual yang memetakan intensitas energi frekuensi suara layaknya sebuah citra gambar. Selain itu, PANNs juga menyediakan arsitektur dengan pendekatan *Hybrid* yang menggabungkan kedua representasi tersebut, yang secara teoritis berpotensi memaksimalkan akurasi deteksi. Ketersediaan variasi ini memunculkan urgensi untuk mengevaluasi arsitektur mana yang paling optimal untuk diterapkan pada kasus ini, apakah berbasis domain waktu, domain frekuensi, atau penggabungan keduanya (*Hybrid*).

Meskipun Kong et al. telah memaparkan tolak ukur kinerja model-model tersebut pada dataset masif AudioSet [7], performa tersebut belum tentu sebanding ketika diterapkan pada kasus penerapan spesifik dengan ketersediaan data yang terbatas (*data scarcity*) seperti pada kasus klasifikasi suara lingkungan perkotaan. Perbedaan karakteristik data ini memunculkan dugaan bahwa kompleksitas arsitektur model *Hybrid* dan *Log-mel Spectrogram* justru memiliki risiko *overfitting* yang lebih tinggi dibandingkan model *Raw Waveform* ketika dilatih pada dataset yang kecil. Ketidakpastian inilah yang menjadi celah penelitian (*research gap*) yang belum terjamah. Oleh karena itu, penelitian ini menjadi krusial untuk mengevaluasi ulang adaptabilitas dan melakukan analisis komparasi ketiga arsitektur tersebut secara spesifik pada dataset UrbanSound8K.

Guna menjawab tantangan adaptabilitas pada dataset terbatas tersebut, penelitian ini bertujuan utama secara teknis untuk menginvestigasi dan membandingkan kinerja tiga pendekatan representasi input, yaitu *Raw Waveform*,

*Log-mel Spectrogram*, dan *Hybrid*. Studi komparasi ini diposisikan sebagai langkah fundamental untuk menemukan konfigurasi model yang paling *robust* (tahan uji) terhadap minimnya data, sekaligus meminimalisir kesalahan deteksi fatal. Dengan demikian, hasil evaluasi ini diharapkan dapat menjadi landasan teknis yang valid bagi pengembangan teknologi asistif yang benar-benar andal untuk menjamin keselamatan 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*, *Log-mel 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*, *Log-mel Spectrogram*, dan *Hybrid*) terhadap performa model *Pre-trained Audio Neural Networks* (PANNs) dalam mengklasifikasikan suara bahaya pada kondisi ketersediaan data yang terbatas.
2. Mengevaluasi 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. Penulis tidak mengumpulkan dataset secara langsung (manual).
2. Lingkup klasifikasi dibatasi pada kategori suara lingkungan yang merepresentasikan indikator bahaya atau peringatan bagi keselamatan fisik di jalan raya.
3. Fokus penelitian terbatas pada eksperimen pelatihan (*training*) dan evaluasi performa model *Deep Learning*, serta tidak mencakup perancangan perangkat keras (*hardware*), pengembangan antarmuka pengguna (*User Interface*), maupun implementasi sistem secara *real-time*.

#### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Memberikan bukti nyata terkait efektivitas metode *Transfer Learning* pada arsitektur PANNs serta perbandingan performa antara representasi input *Raw Waveform*, *Log-mel Spectrogram*, dan *Hybrid* dalam mengatasi keterbatasan 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 ?? dan digunakan sebagai referensi berdasarkan penelitian yang berkaitan dengan penelitian ini.

#### **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**

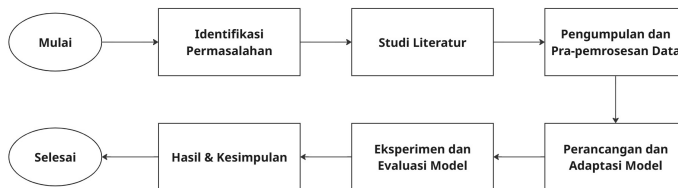
Korean Ballad adalah genre musik asal Korea yang dikenal lembut, emosional, dan memiliki aransemen instrumental kuat.

## BAB III

### METODE PENELITIAN

#### 3.1 Alur Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan eksperimen sistematis untuk mengevaluasi kinerja adaptasi arsitektur *Pre-trained Audio Neural Networks* (PANNs). Pendekatan penelitian didasarkan pada metode *Transfer Learning* dengan desain studi komparatif terhadap variasi representasi input audio. Secara rinci, kerangka kerja dan alur penelitian digambarkan pada diagram alur (*flowchart*) dalam Gambar 3.1 berikut.



Gambar 3.1 Alur Penelitian

#### 3.2 Penjabaran Langkah Penelitian

Berdasarkan diagram alur penelitian yang ditunjukkan pada gambar 3.1, penjelasan rinci terkait langkah-langkah penelitian yang dilakukan akan dijabarkan di bawah ini.

##### 3.2.1 Identifikasi Permasalahan

Tahap awal penelitian dimulai dengan mengidentifikasi urgensi pengembangan sistem alat bantu bagi penyandang tunarungu yang berorientasi pada standar keselamatan tinggi (*safety-critical*). Alat bantu ini diharapkan tidak hanya mengamplifikasi sinyal, tetapi juga mampu mengenali suara bahaya

di lingkungan sekitar secara otomatis. Tantangan teknis utama yang dihadapi adalah keterbatasan jumlah sampel pada dataset target (*limited dataset*) yang spesifik untuk suara bahaya di lingkungan. Melatih model *Deep Learning* dari awal (*from scratch*) dengan data terbatas berisiko tinggi menyebabkan *overfitting* dan performa yang buruk. Oleh karena itu, penggunaan arsitektur *Pre-trained Audio Neural Networks* (PANNs) yang telah dilatih pada dataset raksasa AudioSet menjadi solusi strategis (*Transfer Learning*) untuk mengatasi kelangkaan data tersebut. Akan tetapi, penerapan strategi ini memunculkan permasalahan baru, yaitu belum diketahuinya representasi input audio manakah antara *Raw Waveform* 1D, *Spectrogram* 2D, atau *Hybrid* yang paling optimal untuk mengadaptasi model PANNs tersebut. Ketidaktepatan pemilihan representasi input pada proses transfer pengetahuan ini berisiko menghambat performa model dalam mengenali bahaya secara akurat.

### 3.2.2 Studi Literatur

Tahap ini dilakukan untuk membangun landasan teoretis yang kokoh sebelum perancangan sistem dimulai. Studi literatur difokuskan pada analisis mendalam terhadap arsitektur *Pre-trained Audio Neural Networks* (PANNs), khususnya varian CNN14, serta mekanisme *Transfer Learning* untuk adaptasi domain. Selain itu, kajian juga dipusatkan pada karakteristik teknis dataset *UrbanSound8K*, terutama terkait struktur metadata dan potensi kebocoran data (*data leakage*). Pemahaman terhadap aspek-aspek tersebut menjadi acuan krusial dalam merancang strategi pembagian data (*splitting strategy*) yang valid dan skenario komparasi representasi input yang objektif.

### 3.2.3 Pengumpulan dan Pra-pemrosesan Data

Data penelitian bersumber dari dataset publik *UrbanSound8K* yang memuat 8732 potongan audio (*slices*) berdurasi kurang dari 4 detik. Tahap pra-pemrosesan dilakukan melalui tiga langkah strategis untuk menjamin integritas



data:

1. **Penyaringan Kelas (*Class Filtering*):** Dari 10 kelas yang tersedia, dilakukan seleksi data untuk mengambil 4 kelas prioritas yang relevan dengan konteks keselamatan tunarungu, yaitu: *car\_horn* (klakson), *siren* (sirine), *gun\_shot* (tembakan), dan *dog\_bark* (gonggongan).
2. **Strategi Partisi Anti-Bocor (*Leakage-Free Splitting*):** Dataset asli disusun berdasarkan metadata *fsID* (sumber rekaman) ke dalam 10 *fold*. Untuk mencegah kebocoran data—dimana potongan audio dari sumber rekaman yang sama masuk ke data latih dan data uji sekaligus—penelitian ini menerapkan strategi pemetaan *fold* (*Fold Mapping*). Ke-10 *fold* asli digabungkan dan dipetakan ulang menjadi 5 *fold* eksperimen (Fold 1-2 menjadi Fold Baru 1, dst.) guna menjaga independensi data uji.
3. **Penyeimbangan Kelas (*Class Balancing*):** Mengingat distribusi data mentah yang tidak seimbang (*imbalanced*), diterapkan teknik *Random Under-sampling* pada setiap *fold* baru. Jumlah sampel pada kelas mayoritas dipangkas secara acak hingga setara dengan jumlah sampel pada kelas minoritas. Langkah ini krusial untuk mencegah model mengalami bias prediksi ke arah kelas yang dominan.

Hasil akhir dari proses ini disimpan dalam format konfigurasi statis (*split.json*) yang menjadi acuan baku bagi seluruh skenario pengujian model.

### 3.2.4 Perancangan dan Adaptasi Model

Penelitian ini menggunakan arsitektur dasar *Cnn14* dari pustaka PANNs yang telah dilatih sebelumnya (*pre-trained*) pada dataset AudioSet. Untuk mengadaptasi model tersebut ke dalam domain permasalahan yang baru, dilakukan serangkaian modifikasi struktural sebagai berikut:

1. **Modifikasi Lapisan Luaran (*Output Layer*):** Lapisan klasifikasi asli yang memiliki 527 *node* (sesuai kelas AudioSet) digantikan dengan

lapisan *Fully Connected* baru yang hanya memiliki 4 *node*. Lapisan ini merepresentasikan probabilitas kelas target: klakson, sirine, tembakan, dan gonggongan. Bobot pada lapisan awal dipertahankan dari model *pre-trained* (*Transfer Learning*), sedangkan lapisan baru diinisialisasi ulang untuk dipelajari selama proses pelatihan.

2. **Perancangan Skenario Input:** Untuk menjawab rumusan masalah mengenai representasi fitur terbaik, dirancang tiga skenario variasi input yang akan diaplikasikan secara terpisah pada arsitektur model:

- **Skenario 1 (1D):** Menggunakan input *Raw Waveform* (gelombang amplitudo terhadap waktu) untuk mengekstraksi fitur langsung dari domain waktu.
- **Skenario 2 (2D):** Menggunakan input *Log-mel Spectrogram* yang diekstraksi secara *on-the-fly* untuk menangkap pola representasi frekuensi-waktu.
- **Skenario 3 (Hybrid):** Menggunakan arsitektur *Wavegram-Logmel* yang menggabungkan fitur domain waktu dan frekuensi secara simultan.

Ketiga variasi model ini akan dilatih menggunakan konfigurasi dan data yang identik untuk memastikan validitas hasil komparasi.

### 3.2.5 Eksperimen dan Evaluasi Model

Tahap akhir melibatkan pelaksanaan pelatihan dan pengujian model secara sistematis menggunakan skema validasi silang 5 lipatan (*5-Fold Cross-Validation*). Setiap skenario input (1D, 2D, dan Hybrid) akan dilatih sebanyak 5 putaran, di mana pada setiap putaran model menggunakan kombinasi *fold* latih dan *fold* uji yang berbeda sesuai konfigurasi yang telah ditetapkan pada tahap pra-pemrosesan.

Evaluasi kinerja model tidak hanya didasarkan pada metrik Akurasi (*Accuracy*) semata, tetapi juga dianalisis secara mendalam menggunakan

*Confusion Matrix*. Penggunaan *Confusion Matrix* sangat krusial dalam konteks sistem keselamatan untuk memetakan distribusi kesalahan prediksi—misalnya, seberapa sering suara sirine salah diklasifikasikan sebagai suara latar belakang atau kelas lainnya.

Hasil evaluasi dari kelima *fold* kemudian dirata-rata untuk mendapatkan skor performa final. Berdasarkan data tersebut, dilakukan analisis komparatif untuk menyimpulkan representasi input manakah yang paling andal dan konsisten untuk diimplementasikan pada sistem alat bantu tunarungu.

### **3.2.6 Analisis Hasil dan Penarikan Kesimpulan**

Tahap penutup penelitian difokuskan pada interpretasi mendalam terhadap data kuantitatif yang diperoleh dari tahap evaluasi. Analisis dilakukan secara komparatif untuk menelaah kekuatan dan kelemahan dari masing-masing representasi input (*Raw Waveform* vs *Spectrogram* vs *Hybrid*) dalam mendeteksi pola suara bahaya.

Berdasarkan analisis tersebut, ditarik kesimpulan akhir mengenai model mana yang memiliki performa paling optimal. Kesimpulan ini tidak hanya menjawab rumusan masalah secara statistik, tetapi juga memberikan rekomendasi teknis mengenai kelayakan implementasi model terpilih pada sistem alat bantu keselamatan bagi penyandang tunarungu.

## **3.3 Alat dan Bahan Tugas Akhir**

Penelitian ini membutuhkan dukungan perangkat keras dan perangkat lunak untuk proses pengembangan, pelatihan model, hingga evaluasi. Selain itu, ketersediaan bahan berupa data dan model dasar juga menjadi komponen krusial.

### **3.3.1 Alat**

Alat yang digunakan dalam penelitian ini dikategorikan menjadi perangkat keras (*hardware*) dan perangkat lunak (*software*):

### 1. Perangkat Keras:

- Komputer/Laptop: Digunakan untuk tahap penulisan kode, pra-pemrosesan data ringan, dan penyusunan laporan (Spesifikasi: RAM 16GB).
- *Cloud Computing Environment*: Google Colab (Pro/Free) yang menyediakan akselerasi GPU (misal: NVIDIA T4) untuk mempercepat proses pelatihan model *Deep Learning*.

### 2. Perangkat Lunak:

- *Integrated Development Environment (IDE)*: Visual Studio Code.
- Bahasa Pemrograman: Python 3.10.
- *Version Control System*: Git dan GitHub.
- Pustaka (*Libraries*):
  - **PyTorch**: *Framework* utama untuk pembangunan dan pelatihan model PANNs.
  - **Librosa**: Untuk pemrosesan audio dan ekstraksi fitur (spektrogram).
  - **Pandas & NumPy**: Untuk manipulasi data matriks dan manajemen *dataframe*.
  - **Scikit-learn**: Untuk evaluasi model (*Confusion Matrix*, *Accuracy*).
  - **Matplotlib/Seaborn**: Untuk visualisasi data.

### 3.3.2 Bahan

Bahan penelitian merujuk pada objek data digital yang diolah dan digunakan dalam eksperimen, meliputi:

1. **Dataset UrbanSound8K**: Kumpulan data audio lingkungan yang terdiri dari 8732 potongan suara (*slices*) yang terbagi dalam 10 kelas, beserta file metadata (.csv) yang memuat informasi *fold* dan *class ID*.
2. **Pre-trained Weights (Cnn14.pth)**: File bobot model PANNs (CNN14)

yang telah dilatih sebelumnya pada dataset AudioSet. File ini digunakan sebagai inisialisasi awal dalam proses *Transfer Learning*.

### 3.4 Metode Pengembangan

Berbeda dengan pengembangan perangkat lunak konvensional, penelitian ini menerapkan metode pengembangan berbasis data (*Data-Driven Development*) dengan pendekatan eksperimental. Inti dari metode ini adalah adaptasi pengetahuan (*Transfer Learning*) dari model pra-latih untuk menyelesaikan permasalahan pada domain baru dengan dataset terbatas.

Metode pengembangan yang digunakan mencakup empat tahapan utama: Akuisisi Data, Rekayasa Data (*Data Engineering*), Adaptasi Model, dan Evaluasi Kinerja.

#### 3.4.1 Prosedur Eksperimen

Seluruh rangkaian eksperimen dijalankan mengikuti alur kerja (*pipeline*) *Deep Learning* standar:

1. **Setup Lingkungan:** Mengonfigurasi *GPU Runtime* pada Google Colab dan instalasi pustaka PANNs.
2. **Pra-pemrosesan:** Memuat file `split.json` untuk membagi data latih dan uji secara otomatis.
3. **Pelatihan (*Training Loop*):** Melatih model selama sejumlah *epoch* tertentu dengan memantau penurunan *loss* dan peningkatan akurasi.
4. **Inferensi:** Menguji model final pada data validasi untuk mendapatkan metrik performa.

#### 3.4.2 Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini tidak dilakukan secara manual (perekaman langsung), melainkan melalui akuisisi dataset sekunder yang telah terstandarisasi.

1. **Sumber Data:** Dataset *UrbanSound8K* diunduh dari repositori resmi

atau sumber terpercaya (misalnya Zenodo atau Kaggle).

2. **Validasi Integritas:** Memastikan struktur folder (*fold1* s.d *fold10*) dan file metadata (*UrbanSound8K.csv*) lengkap dan tidak korup.
3. **Penyaringan (*Filtering*):** Mengambil hanya sampel data yang memiliki *classID* sesuai target: *car\_horn* (1), *dog\_bark* (3), *gun\_shot* (6), dan *siren* (8).

### 3.4.3 Strategi Transfer Learning

Sebagai pengganti metode pengembangan sistem konvensional, penelitian ini menerapkan strategi *Transfer Learning* dengan langkah-langkah:

1. **Inisialisasi Bobot:** Memuat bobot model CNN14 yang telah dilatih pada *AudioSet*.
2. **Pembekuan Sebagian (*Partial Freezing*):** Membekukan bobot pada lapisan ekstraksi fitur awal (*lower layers*) agar tidak rusak saat pelatihan awal, dan hanya melatih lapisan-lapisan akhir.
3. **Fine-Tuning:** Melakukan pelatihan ulang pada model dengan *learning rate* yang kecil untuk menyesuaikan representasi fitur model dengan karakteristik suara lingkungan pada dataset target.

### 3.4.4 Metode Pengujian Penelitian

Untuk menghindari bias yang sering terjadi pada metode *Hold-Out* (80:20) biasa, penelitian ini menggunakan metode pengujian yang lebih ketat:

1. **Stratified Group 5-Fold Cross-Validation:** Data dibagi menjadi 5 bagian (*folds*). Pada setiap iterasi, 4 bagian digunakan untuk melatih model, dan 1 bagian digunakan untuk pengujian. Proses ini diulang 5 kali hingga semua bagian pernah menjadi data uji. Metode ini menjamin:
  - Distribusi kelas seimbang di setiap lipatan (*Stratified*).
  - Tidak ada kebocoran data dari sumber rekaman yang sama (*Grouped by fsID*).

## 2. Metrik Evaluasi:

- **Akurasi (Accuracy):** Untuk mengukur performa global model.
- **Confusion Matrix:** Untuk menganalisis detail kesalahan prediksi antar kelas (misal: Sirine tertukar dengan Klakson).

### 3.5 Ilustrasi Perhitungan Metode

Bagian ini menyajikan simulasi proses kerja sistem, mulai dari representasi data input, interpretasi keluaran model (*model output*), hingga perhitungan metrik evaluasi. Ilustrasi ini menggunakan sampel data dummy untuk mempermudah pemahaman logika sistem.

#### 3.5.1 Ilustrasi Input dan Output Model

Misalkan terdapat sebuah file audio input  $x$  yang berisi suara "Sirine".

1. **Input:** Audio tersebut diubah menjadi *Log-mel Spectrogram*. Model menerima input berupa tensor (matriks) 2D.
2. **Proses:** Model PANNs memproses input tersebut melewati lapisan konvolusi.
3. **Output Probabilitas:** Karena sistem memiliki 4 kelas target, lapisan terakhir model akan mengeluarkan vektor probabilitas ( $P$ ) untuk setiap kelas:
  - Index 0: *Car Horn*
  - Index 1: *Dog Bark*
  - Index 2: *Gun Shot*
  - Index 3: *Siren*

Contoh keluaran model (*Softmax Output*):

$$P = [0.05, 0.10, 0.05, 0.80]$$

4. **Keputusan Kelas:** Sistem mengambil nilai probabilitas tertinggi ( $\text{Argmax}$ ).  
 Karena prediksi (*Siren*) sama dengan label asli (*Siren*), maka prediksi dinyatakan **BENAR**.

3.5.2 Ilustrasi Perhitungan Evaluasi

Untuk mengukur kinerja model, digunakan metrik Akurasi. Berikut adalah ilustrasi perhitungan manual menggunakan 10 sampel data uji fiktif dengan 4 kelas.

Tabel 3.1 Contoh Data Uji dan Hasil Prediksi

No	Label Asli	Prediksi Model	Status
1	Siren	Siren	Benar
2	Car Horn	Car Horn	Benar
3	Dog Bark	Dog Bark	Benar
4	Gun Shot	Gun Shot	Benar
5	Siren	<b>Car Horn</b>	<b>Salah</b>
6	Dog Bark	Dog Bark	Benar
7	Car Horn	Car Horn	Benar
8	Gun Shot	Gun Shot	Benar
9	Siren	Siren	Benar
10	Dog Bark	<b>Siren</b>	<b>Salah</b>

**Perhitungan Akurasi:** Berdasarkan Tabel simulasi di atas, diketahui:

- Jumlah Data Total ( $N$ ) = 10
- Jumlah Prediksi Benar ( $TP + TN$ ) = 8
- Jumlah Prediksi Salah = 2

Maka akurasi dihitung menggunakan rumus  $Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data} \times 100\%$  (Rumus 3.1)  
Dengan demikian, data uji ini memiliki tingkat akurasi sebesar 80%. (Rumus 3.2)

Prinsip perhitungan yang sama akan diterapkan secara otomatis oleh komputer pada ribuan data uji menggunakan *Confusion Matrix* selama proses validasi.

3.6 Rancangan Pengujian dan Hipotesis

Rancangan pengujian disusun untuk memverifikasi efektivitas model dalam menyelesaikan masalah klasifikasi suara lingkungan. Pengujian ini difokuskan pada dua aspek utama:

1. **Pengujian Komparatif (Comparative Testing):** Menguji tiga skenario



model (Input 1D, 2D, dan Hybrid) secara terpisah menggunakan data uji yang identik untuk melihat model mana yang paling unggul dalam menangani dataset terbatas.

2. **Pengujian Kinerja Klasifikasi:** Mengukur kemampuan model dalam mengenali 4 kelas target (klakson, sirine, tembakan, gonggongan) dengan melihat tingkat kesalahan prediksi antar kelas melalui *Confusion Matrix*.

### **Hipotesis Penelitian**

Berdasarkan landasan teori, hipotesis awal yang diajukan dalam penelitian ini adalah:

- H0: Tidak terdapat perbedaan performa yang signifikan antara penggunaan input *Raw Waveform*, *Spectrogram*, maupun *Hybrid* pada arsitektur PANNs.
- H1: Representasi input **Hybrid (Wavegram-Logmel)** diharapkan menghasilkan akurasi tertinggi dibandingkan input tunggal (1D atau 2D), karena kemampuannya menangkap informasi fitur dari domain waktu dan frekuensi secara simultan, yang krusial untuk membedakan suara lingkungan yang kompleks.

## DAFTAR PUSTAKA

- [1] World Health Organization. *Global Status Report on Road Safety 2023*. Geneva: World Health Organization, 2023. ISBN: 9789240086517.
- [2] Birgitta Thorslund et al. “Effects of hearing loss on traffic safety and mobility”. *European Transport Research Review* 5 (2013), pp. 113–121.
- [3] 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.
- [4] Michael Büchler et al. “Sound classification in hearing aids inspired by auditory scene analysis”. *The Journal of the Acoustical Society of America* 118.3 (2005), pp. 2057–2057.
- [5] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. “A Dataset and Taxonomy for Urban Sound Research”. *Proceedings of the 22nd ACM International Conference on Multimedia*. ACM. 2014, pp. 1041–1044.
- [6] Karol J Piczak. “Environmental sound classification with convolutional neural networks”. *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 2015, pp. 1–6.
- [7] 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.