

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

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	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
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
BAB III METODE PENELITIAN	8
3.1 Alur Penelitian	8
3.2 Penjabaran Langkah Penelitian	8
3.2.1 Identifikasi Permasalahan	9

3.2.2	Studi Literatur	9
3.2.3	Pengumpulan dan Pra-pemrosesan Data	10
3.2.4	Adaptasi Model	11
3.2.5	Evaluasi Model	11
3.2.6	Hasil dan Kesimpulan	11
3.3	Alat dan Bahan Tugas Akhir.....	12
3.3.1	Alat.....	12
3.3.2	Bahan.....	12
3.4	Metode Pengembangan	13
3.4.1	Model Pengembangan	13
3.4.2	Prosedur Pengembangan	13
3.4.3	Cara Pengumpulan Data	14
3.4.4	Strategi <i>Transfer Learning</i>	14
3.5	Ilustrasi Perhitungan Metode	14
3.5.1	Matriks Kebingungan (Confusion Matrix)	15
3.5.2	Perhitungan Metrik	15
3.6	Rancangan Pengujian	16
3.6.1	Lingkungan Pengujian.....	16
3.6.2	Pengujian Fungsional (Kinerja Klasifikasi).....	16
3.6.3	Pengujian Non-Fungsional (Efisiensi Sistem).....	17
DAFTAR PUSTAKA	18

DAFTAR TABEL

Tabel 3.1 Contoh Matriks Kebingungan Multikelas (Data Hipotetik). 15

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 gongongan anjing yang mengejar atau klakson kendaraan yang melaju kencang dari arah belakang, sering kali 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 mengantikan 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* dalam mengklasifikasikan suara tanda bahaya yang mengancam keselamatan penyandang Tuna Rungu. 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 pengembangan 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

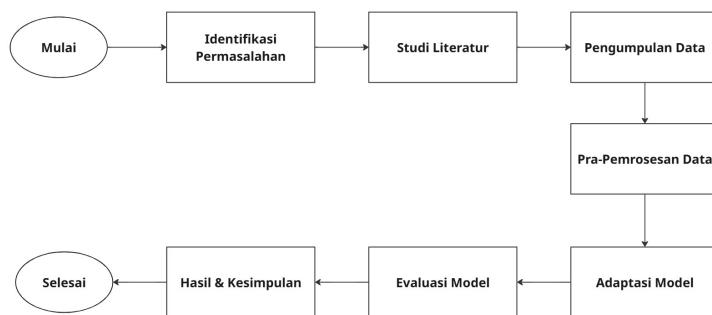
Korean Ballad adalah genre musik asal Korea yang dikenal lembut, emosional, dan memiliki aransemen instrumental kuat.

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. 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 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 pra-pemrosesan data dan adaptasi model dimulai. Studi literatur difokuskan pada analisis mendalam terhadap arsitektur *Pre-trained Audio Neural Networks* (PANNs), dengan perwakilan tiga model yang masing-masing memiliki representasi input yang berbeda, 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. Kajian literatur juga mencakup pemilihan metode evaluasi yang relevan untuk kasus klasifikasi multikelas pada dataset tidak seimbang, guna memastikan hasil eksperimen dapat diukur validitasnya secara objektif.

3.2.3 Pengumpulan dan Pra-pemrosesan Data

Data penelitian bersumber dari dataset publik *UrbanSound8K*. 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* (gongongan).
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 berupa `split.json` yang menjadi acuan baku bagi seluruh skenario pengujian model.

3.2.4 Adaptasi Model

Penelitian ini menggunakan tiga arsitektur dasar terbaik dari setiap representasi input nya, yaitu *Res1dNet31*, *ResNet38*, dan *Wavegram-Logmel-CNN* dari pustaka PANNs yang telah dilatih sebelumnya (*pre-trained*) pada dataset AudioSet. Untuk mengadaptasi model tersebut ke dalam domain permasalahan yang baru, dilakukan serangkaian langkah modifikasi pada arsitektur yang akan digunakan. Untuk langkah adaptasi yang lebih rinci, hal tersebut dijelaskan pada bagian 3.4.

3.2.5 Evaluasi Model

Setelah model diadaptasi, maka dilakukan evaluasi performa model menggunakan pendekatan *Multi-metric Evaluation*. 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 Hasil dan 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 Tuna Rungu.

3.3 Alat dan Bahan Tugas Akhir

Penelitian ini membutuhkan dukungan alat untuk proses adaptasi, 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 adalah:

1. *Visual Studio Code* sebagai *text editor*.
2. Kaggle Notebook dengan *GPU Runtime* sebagai lingkungan pengembangan dan eksekusi kode.
3. Python versi 3.10.16

3.3.2 Bahan

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

1. Dataset UrbanSound8K yang merupakan 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*, yaitu file bobot model PANNs, yaitu *Res1dNet31.pth*, *ResNet38.pth*, dan *Wavegram_Logmel_CNN.pth*, yang telah dilatih sebelumnya pada dataset AudioSet. File ini digunakan sebagai inisialisasi awal dalam proses *Transfer Learning*.
3. Dokumen Referensi Utama yang berupa paper ilmiah ”PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition” yang digunakan sebagai acuan standar arsitektur model [7].

3.4 Metode Pengembangan

Bagian ini menjelaskan pendekatan teknis yang digunakan untuk membangun sistem klasifikasi suara. Berbeda dengan pengembangan perangkat lunak konvensional, penelitian ini berfokus pada pengembangan model cerdas berbasis data.

3.4.1 Model Pengembangan

Penelitian ini mengadopsi model pengembangan *Machine Learning Pipeline*. Model ini dipilih karena penelitian bersifat eksperimental dan berorientasi pada data (*data-driven*). Alur pengembangan tidak bersifat linier kaku, melainkan bersifat iteratif terutama pada tahap pelatihan dan evaluasi model untuk mencapai performa terbaik.

3.4.2 Prosedur Pengembangan

Secara spesifik, prosedur pengembangan model dalam penelitian ini terdiri dari empat tahapan utama:

1. **Pra-pemrosesan Data (*Data Preprocessing*):** Tahap ini mengubah data mentah menjadi format yang siap dilatih. Termasuk di dalamnya adalah konversi audio ke representasi visual (*Spectrogram*) untuk input model 2D dan *Hybrid*, serta pembagian data menggunakan skema *5-Fold Cross Validation*.
2. **Adaptasi Arsitektur (*Model Adaptation*):** Tahap memodifikasi arsitektur PANNs agar sesuai dengan 4 kelas target. Lapisan klasifikasi asli (527 *nodes*) diganti dengan lapisan *Fully Connected* baru (4 *nodes*).
3. **Pelatihan Model (*Model Training*):** Tahap pembelajaran mesin menggunakan strategi *Transfer Learning*. Bobot model dilatih ulang (*fine-tuning*) menggunakan optimasi *AdamW* dan fungsi kerugian *Cross-Entropy Loss*.
4. **Evaluasi (*Model Evaluation*):** Tahap pengujian kinerja model

menggunakan data validasi yang belum pernah dilihat model sebelumnya untuk mengukur kemampuan generalisasi.

3.4.3 Cara Pengumpulan Data

Sesuai dengan karakteristik penelitian *Deep Learning* yang membutuhkan data dalam jumlah besar, metode pengumpulan data yang digunakan adalah Studi Dokumentasi Dataset Sekunder.

1. **Sumber Data:** Data tidak dikumpulkan melalui perekaman langsung (primer), melainkan diunduh dari repositori publik *UrbanSound8K*.
2. **Teknik Sampling:** Pengambilan sampel dilakukan dengan teknik *Purposive Sampling*, yaitu hanya mengambil data yang memenuhi kriteria inklusi: (a) Termasuk dalam 4 kelas bahaya (Klakson, Sirine, Tembakan, Gonggongan), dan (b) Memiliki durasi audio yang valid (≥ 4 detik).

3.4.4 Strategi Transfer Learning

Untuk mengatasi keterbatasan jumlah data pada dataset target, pengembangan model menerapkan teknik *Partial Freezing*:

- **Frozen Layers:** Lapisan awal (*feature extractor*) dibekukan bobotnya agar fitur dasar suara yang sudah dipelajari dari AudioSet tidak hilang.
- **Trainable Layers:** Hanya lapisan akhir (*classifier*) dan beberapa blok konvolusi terakhir yang diizinkan untuk memperbarui bobotnya selama proses pelatihan.

3.5 Ilustrasi Perhitungan Metode

Pada bagian ini disajikan ilustrasi perhitungan metrik evaluasi berbasis *Confusion Matrix* untuk kasus klasifikasi multikelas (4 kelas). Data yang digunakan bersifat hipotetik dengan jumlah sampel $N = 20$, bertujuan untuk mengilustrasikan mekanisme perhitungan *Accuracy*, *Precision*, dan *Recall*.

3.5.1 Matriks Kebingungan (Confusion Matrix)

Contoh hasil prediksi model terhadap 20 data uji yang terdiri dari kelas A (Klakson), B (Sirine), C (Tembakan), dan D (Gonggongan) disajikan pada Tabel 3.2.

Tabel 3.1 Contoh Matriks Kebingungan Multikelas (Data Hipotetik)

		Prediksi Model			
		A	B	C	D
Aktual	A (Klakson)	4	1	0	0
	B (Sirine)	1	4	0	0
	C (Tembakan)	0	0	5	0
	D (Gonggongan)	0	1	0	4

Keterangan diagonal utama (cetak tebal) menunjukkan prediksi yang benar (*True Positive* untuk masing-masing kelas).

3.5.2 Perhitungan Metrik

Berdasarkan Tabel 3.2, total prediksi benar adalah jumlah diagonal utama: $4 + 4 + 5 + 4 = 17$.

1. Akurasi Global (*Overall Accuracy*)

Mengukur rasio prediksi benar terhadap total sampel $\frac{\sum \text{Diagonal}}{N} = \frac{17}{20} = 85\%$ (Rumus 3.1)

2. Perhitungan Per-Kelas (Contoh: Kelas B/Sirine)

Dalam klasifikasi multikelas, *Precision* dan *Recall* dihitung secara terpisah untuk setiap kelas (*One-vs-Rest*). Berikut contoh perhitungan untuk **Kelas B (Sirine)**:

- **True Positive (TP):** Aktual Sirine diprediksi Sirine = 4.
- **False Positive (FP):** Aktual BUKAN Sirine, tapi diprediksi Sirine (Lihat Kolom B selain diagonal: 1 dari A + 1 dari D) = 2.
- **False Negative (FN):** Aktual Sirine, tapi diprediksi BUKAN Sirine (Lihat Baris B selain diagonal: 1 ke A) = 1.

Maka metrik untuk Kelas Sirine adalah:

$$\text{Precision}_B = \frac{TP}{TP+FP} = \frac{4}{4+2} = \frac{4}{6} \approx 66,67\% \quad (\text{Rumus 3.2})$$

Nilai akhir sistem (*Macro Average*) diperoleh dengan merata-ratakan nilai metrik dari keempat kelas tersebut.

3.6 Rancangan Pengujian

Rancangan pengujian disusun untuk memverifikasi apakah sistem yang dikembangkan memenuhi spesifikasi kebutuhan, baik dari segi kemampuan deteksi (fungsional) maupun efisiensi komputasi (non-fungsional). Pengujian dilakukan secara komparatif terhadap ketiga variasi model yang diajukan.

3.6.1 Lingkungan Pengujian

Seluruh skenario pengujian dijalankan pada lingkungan yang seragam untuk menjamin validitas perbandingan (*Apple-to-Apple Comparison*). Spesifikasi lingkungan pengujian adalah sebagai berikut:

1. **Perangkat Keras:** Pengujian inferensi dilakukan menggunakan *GPU Runtime* NVIDIA T4 pada platform Kaggle.
2. **Perangkat Lunak:** Kerangka kerja *PyTorch* dengan versi pustaka yang dikunci (*fixed version*) untuk menghindari perbedaan optimasi pustaka.

3.6.2 Pengujian Fungsional (Kinerja Klasifikasi)

Pengujian fungsional bertujuan untuk mengukur kehandalan model dalam menjalankan fungsi utamanya, yaitu mengklasifikasikan audio input ke dalam 4 kelas target secara akurat.

1. **Metode:** Menggunakan *Stratified 5-Fold Cross-Validation* untuk mendapatkan nilai performa yang objektif dan tidak bias oleh pembagian data tertentu.
2. **Parameter Ukur:**
 - **Macro F1-Score:** Sebagai metrik utama untuk melihat kinerja rata-rata yang adil pada dataset yang tidak seimbang.
 - **Recall (Sensitivitas):** Khusus dipantau untuk kelas bahaya (Sirine dan Tembakan) guna memastikan aspek keselamatan pengguna.

- **Confusion Matrix:** Untuk memvalidasi apakah model mampu membedakan antar kelas yang mirip dengan benar.
3. **Kriteria Keberhasilan:** Model dinyatakan lolos uji fungsional jika mampu menghasilkan nilai *F1-Score* di atas garis dasar (*baseline*) model klasifikasi sederhana atau penelitian terdahulu yang relevan.

3.6.3 Pengujian Non-Fungsional (Efisiensi Sistem)

Mengingat tujuan akhir sistem adalah sebagai alat bantu dengar yang bekerja secara *real-time*, aspek performa non-fungsional menjadi krusial.

1. **Waktu Inferensi (Inference Time):** Mengukur rata-rata waktu yang dibutuhkan model untuk memproses satu sampel audio berdurasi 4 detik, mulai dari data masuk hingga keluar prediksi kelas. Pengukuran dilakukan dalam satuan milidetik (ms). Semakin kecil waktu inferensi, semakin layak model diterapkan pada sistem *real-time*.
2. **Ukuran Model (Model Size):** Mengukur besaran file bobot model (.pth) yang dihasilkan setelah pelatihan. Parameter ini menguji efisiensi penyimpanan (*storage*) jika nantinya model akan ditanamkan (*embedded*) pada perangkat keras dengan memori terbatas.

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.