ANALISIS PERBANDINGAN PERFORMA ARSITEKTUR PANNS DENGAN PENDEKATAN INPUT BERBASIS *RAW WAVEFORM*, *LOG-MEL SPECTROGRAM*, DAN *HYBRID* UNTUK KLASIFIKASI SUARA BAHAYA

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 TABEL

DAFTAR GAMBAR

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 dari 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. [7] telah memaparkan tolak ukur kinerja model-model tersebut pada dataset masif AudioSet, 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 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 penelitan 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

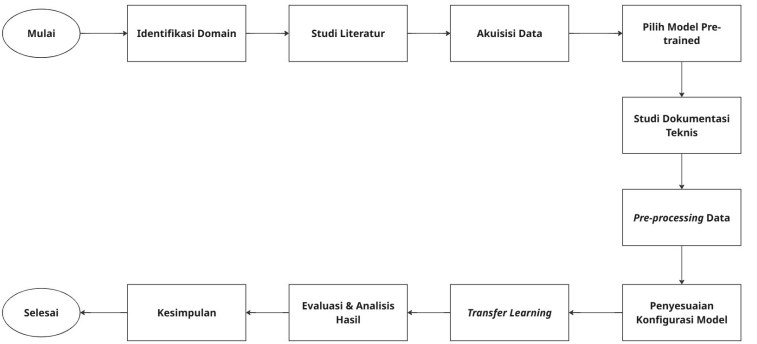
Bab ini membahas kesimpulan dari hasil penelitian dan juga saran untuk penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

BAB III
METODE PENELITIAN

3.1 Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis guna memastikan model klasifikasi suara yang dikembangkan dapat bekerja optimal pada domain keselamatan penyandang Tuna Rungu. Diagram alur penelitian ditunjukkan pada Gambar 3.1.



Gambar 3.1 Diagram Alur Penelitian

Penjelasan rinci mengenai tahapan penelitian adalah sebagai berikut:

1. **Identifikasi Domain Penelitian**

   Tahap awal dilakukan dengan mencari referensi domain penelitian untuk klasifikasi suara. Berdasarkan pertimbangan ketersediaan dataset dan urgensi model klasifikasi, domain yang dipilih adalah domain keselamatan publik yang difokuskan untuk alat bantu penyandang Tuna Rungu.

2. **Studi Literatur**

   Dilakukan kajian literatur mendalam untuk menemukan urgensi penelitian, khususnya mengenai kebutuhan teknologi asistif yang mampu mengurangi

risiko kecelakaan bagi penyandang Tuna Rungu melalui pengenalan sinyal bahaya.

3. **Identifikasi dan Akuisisi Dataset**

   Akuisisi dataset dilakukan dengan mengambil dataset sekunder yang sudah terstandarisasi, yaitu UrbanSound8K. Akuisisi dataset primer tidak dilakukan demi menghindari *device bias* akibat perekaman dataset dengan perangkat non-standar (seperti smartphone) serta kendala regulasi keamanan pada perekaman kelas *gun_shot*. Maka dari itu, dataset sekunder terstandarisasi menjadi solusi yang valid dalam penelitian ini. Dari dataset tersebut, dilakukan seleksi kelas sinyal bahaya dengan strategi antisipasi terhadap kendala ketidakseimbangan data (*imbalanced data*) yang telah disiapkan sejak awal.

4. **Pemilihan Model *Pre-trained***

   Mengingat keterbatasan data berisiko menyebabkan kegagalan pelatihan *from scratch*, digunakan pendekatan *Transfer Learning* memanfaatkan model *Pre-trained*. Arsitektur PANNs (*Pre-trained Audio Neural Networks*) dipilih karena telah dilatih pada dataset masif dan memiliki performa yang teruji.

5. **Studi Dokumentasi Teknis**

   Tahap ini mempelajari karakteristik dataset dan model yang digunakan. Fokus utama dalam tahap ini adalah memahami mekanisme pembagian data (fold) pada dataset guna mencegah kebocoran data (*data leakage*) dan mempelajari variasi representasi input pada arsitektur PANNs untuk menentukan skenario komparasi yang tepat.

6. **Pra-pemrosesan Dataset (*Pre-processing*)**

   Serangkaian proses dilakukan untuk mengubah data mentah menjadi format yang siap latih, meliputi penanganan struktur *fold*, seleksi kelas, dan penyesuaian format audio, serta penerapan augmentasi temporal untuk menstandarisasi durasi input.

7. **Penyesuaian Konfigurasi Model**

   Dilakukan modifikasi pada arsitektur model agar sesuai dengan tujuan klasifikasi 4 kelas bahaya, serta penerapan strategi untuk menangani ketidakseimbangan data.

8. **Uji Coba *Transfer Learning***

   Tahapan inti di mana model dilatih untuk mengenali karakteristik suara spesifik. Proses ini melibatkan eksperimen *Trial and Error* untuk menemukan konfigurasi parameter pelatihan yang paling optimal.

9. **Evaluasi dan Analisis Hasil**

   Performa model hasil *Fine-tuning* dievaluasi menggunakan metrik *Confusion Matrix*, *F1-Score*, serta analisis grafik *Loss-Accuracy* dan waktu komputasi. Output model diuji validitasnya untuk memastikan kelayakan implementasi.

10. **Kesimpulan**

    Berdasarkan hasil evaluasi, dilakukan perbandingan komparatif untuk menyimpulkan model mana yang memiliki performa paling unggul dan stabil.

3.2 Akuisisi Data

Sumber data utama dalam penelitian ini adalah dataset publik **UrbanSound8K**. Dataset diunduh secara manual dari repositori Kaggle dalam format terkompresi (`.zip`). Setelah diekstraksi, struktur dataset terdiri dari 10 folder (masing-masing mewakili satu *fold*) beserta satu file metadata (`.csv`) yang memuat informasi nama file audio, *class ID*, dan *fold* asal.

3.3 Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menjamin integritas data dan mencegah kebocoran informasi (*data leakage*) selama proses pelatihan.

3.3.1 Strategi Pembagian Data (*Fold Mapping*)

Dataset UrbanSound8K secara bawaan terbagi ke dalam 10 *fold*. Untuk efisiensi eksperimen tanpa melanggar aturan independensi data, penelitian ini menggabungkan 10 *fold* tersebut menjadi 5 *fold* eksperimen baru. Penggabungan dilakukan secara berurutan (misalnya Fold 1 dan 2 menjadi Fold Baru 1) tanpa pengacakan data antar *fold*.

Setelah penggabungan, dilakukan pembagian set data menjadi Data Latih (*Train*) dan Data Uji (*Test*). Pemisahan ini krusial untuk memastikan model diuji menggunakan data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi mencerminkan kemampuan model mempelajari karakteristik suara, bukan sekadar menghafal data.

3.3.2 Seleksi Kelas (*Class Filtering*)

Tidak seluruh kelas pada dataset UrbanSound8K relevan dengan konteks keselamatan Tuna Rungu. Oleh karena itu, dilakukan penyaringan untuk hanya mengambil 4 kelas prioritas yang merepresentasikan sinyal bahaya, yaitu:

- *gun_shot* (tembakan)
- *siren* (sirine)
- *dog_bark* (gonggongan)
- *car_horn* (klakson)

3.4 Konfigurasi Model

3.4.1 Adaptasi Arsitektur (*Transfer Learning*)

Penyesuaian arsitektur dilakukan pada lapisan keluaran (*output layer*) dan mekanisme pembekuan bobot (*Freeze Base*). Jumlah *neuron* pada lapisan akhir disesuaikan menjadi 4 *node* (sesuai jumlah kelas target). Sementara itu, lapisan ekstraktor fitur (*base model*) dibekukan agar model tidak menghapus ”ingatan” fitur dasar yang telah dipelajari dari dataset besar sebelumnya (AudioSet). Strategi ini memastikan model memiliki inisialisasi bobot yang baik dan hanya

perlu mempelajari pola baru yang spesifik pada dataset target.

3.4.2 Penanganan Ketidakseimbangan Data (*Weight Penalty*)

Untuk mengatasi distribusi data yang tidak seimbang antar kelas, diterapkan mekanisme *Weight Penalty* pada fungsi kerugian (*Loss Function*). Metode ini bekerja dengan memberikan bobot penalti yang lebih besar ketika model salah memprediksi kelas minoritas (jumlah sampel sedikit). Hal ini memaksa model untuk memberikan perhatian yang setara pada semua kelas, mencegah bias prediksi ke arah kelas mayoritas.

3.5 Konfigurasi Parameter Eksperimen

Penelitian ini menggunakan model *Pre-trained Audio Neural Networks* (PANNs) sebagai kerangka dasar (*backbone*). Model ini telah dilatih sebelumnya (*pre-trained*) menggunakan dataset AudioSet yang berskala besar. Untuk mengadaptasi model tersebut ke dalam kasus klasifikasi 4 kelas pada dataset UrbanSound8K, dilakukan metode *Transfer Learning* dengan membekukan (*freeze*) lapisan ekstraksi fitur awal dan hanya melakukan *fine-tuning* pada lapisan *Fully Connected* (FC) terakhir.

Agar hasil evaluasi antar model (*ResNet38*, *Res1dNet31*, dan *Wavegram-Logmel-CNN*) dapat diperbandingkan secara adil (*apple-to-apple*), seluruh proses pelatihan menggunakan konfigurasi *hyperparameter* yang seragam. Rincian parameter konfigurasi yang dikendalikan dalam eksperimen ini disajikan pada Tabel 3.1.

Tabel 3.1 Parameter Konfigurasi Pelatihan

Kategori	**Konfigurasi / Nilai**
Preprocessing Data	
Input Sampling Rate	32.000 Hz
Input Duration	5 Detik (160.000 samples)
Teknik Augmentasi	*Random Cropping* & *Zero-padding*
Skema Validasi	5-Fold Cross Validation (Group Split)
Hyperparameter Training	
Batch Size	8
Epoch	15
Optimizer	Adam
Learning Rate	0.001 ($1e^{-3}$)
Loss Function	Cross Entropy dengan *Class Weights*
Reproducibilitas	
Random Seed	42
Perangkat Keras	GPU NVIDIA GeForce GTX 1050

3.6 Analisis dan Evaluasi

Setelah proses pelatihan selesai, kinerja model diukur menggunakan empat indikator utama:

1. **Confusion Matrix:** Digunakan untuk melihat detail distribusi prediksi benar dan salah pada setiap kelas spesifik.
2. **F1-Score:** Digunakan sebagai metrik utama untuk mengukur presisi dan sensitivitas model secara harmonis, memastikan semua kelas bahaya terdeteksi dengan baik.
3. **Grafik Loss & Accuracy:** Digunakan untuk memantau proses konvergensi model selama pelatihan dan mendeteksi indikasi *overfitting* atau *underfitting*.
4. **Waktu Training:** Digunakan sebagai acuan efisiensi komputasi model.

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.