INTISARI

MODEL KLASIFIKASI AUDIO KENDARAAN PRIORITAS YANG EFISIEN DALAM PENGGUNAAN MEMORI DAN WAKTU INFERENSI

Oleh

Dwi Ahmad Dzulhijjah

23/530908/PPA/06750

Kendaraan prioritas seperti ambulans dan pemadam kebakaran memerlukan prioritas lalu lintas pada keadaan darurat. Sistem deteksi sirine berbasis audio sebelumnya telah menunjukkan akurasi tinggi (≈95%) dengan menggunakan ekstraksi fitur multi-domain (waktu, frekuensi, dan Doppler) serta teknik ensemble. Namun, pendekatan berbasis multi-domain dan stacked ensemble tersebut memiliki kompleksitas komputasi yang tinggi, menghasilkan total *run time inferensi* selama 37,4 ms dan penggunaan memori sebesar 14 MB.

Penelitian ini mengusulkan penggunaan *Mel-Filterbank Energy (MFE)* sebagai metode ekstraksi fitur yang lebih efisien. MFE mengekstrak energi spektral langsung dari deret *mel-filterbank* tanpa memerlukan tahap transformasi *DCT* seperti pada *MFCC*, sehingga menurunkan kompleksitas komputasi sekaligus tetap mempertahankan informasi spektral yang relevan. Dengan menggunakan dataset Emergency Vehicle Siren Sounds yang berisi 600 sampel audio, dikembangkan beberapa pipeline baru yang memanfaatkan kombinasi fitur *Statistical MFE* dan *Raw MFE* untuk model *ensemble*, serta arsitektur 1D-CNN yang ringan. Evaluasi dilakukan menggunakan *5-fold stratified cross-validation* dengan mengukur akurasi dan F1-score klasifikasi, waktu inferensi, serta konsumsi memori baik pada tahap ekstraksi fitur maupun model.

Pipeline usulan berbasis Selected Statistical MFE dengan SVM yang telah dioptimasi mencapai akurasi 97,0% (±1,0%) dengan *run time* total sekitar 4,1 ms per sampel dan penggunaan memori total 1,9 MB. Model 1D-CNN berbasis Statistical MFE mencapai akurasi 93,2% (±1,2%) dengan waktu inferensi sekitar 6,3 ms dan penggunaan memori 13,2 MB. Usulan berbasis MFE mampu meningkatkan akurasi sebesar +2,0% (dari 95,0% menjadi 97,0%), mengurangi waktu inferensi total hingga 89,0% (dari 37,4 ms menjadi 4,1 ms), dan menurunkan penggunaan memori hingga 86,4% (dari 14 MB menjadi 1,9 MB). Hasil hyperparameter tuning menunjukkan peningkatkan performa model baseline sebesar 1,3% dan model usulan sebesar 1,0%.

Kata Kunci: Klasifikasi Audio, Kendaraan Prioritas, *Mel-filterbank Energy*, *Ensemble Learning*, *1D CNN*

ABSTRACT

AN EFFICIENT AUDIO CLASSIFICATION MODEL FOR EMERGENCY VEHICLES IN TERMS OF MEMORY USAGE AND INFERENCE TIME

By

Dwi Ahmad Dzulhijjah

23/530908/PPA/06750

Priority vehicles such as ambulances and fire trucks require traffic priority in emergency situations. Previous audio-based siren detection systems have achieved high accuracy (≈95%) using multi-domain feature extraction (time, frequency, and Doppler) and ensemble techniques. However, such multi-domain and stacked ensemble approaches exhibit high computational complexity, resulting in a total inference run time of 37.4 ms and memory usage of 14 MB.

This study proposes the use of Mel-Filterbank Energy (MFE) as a more efficient feature extraction method. MFE extracts spectral energy directly from the mel-filterbank array without requiring the Discrete Cosine Transform (DCT) step as in MFCC, thereby reducing computational complexity while preserving relevant spectral information. Using the Emergency Vehicle Siren Sounds dataset containing 600 audio samples, several new pipelines were developed that exploit combinations of Statistical MFE and Raw MFE features for ensemble models, as well as lightweight 1D-CNN architectures. The evaluation was conducted using 5-fold stratified cross-validation, measuring classification accuracy and F1-score, inference time, and memory consumption for both the feature extraction and model stages.

The proposed pipeline based on Selected Statistical MFE with optimized SVM achieves an accuracy of 97.0% (±1.0%) with a total run time of approximately 4.1 ms per sample and total memory usage of 1.9 MB. The 1D-CNN model based on Statistical MFE attains an accuracy of 93.2% (±1.2%) with an inference time of approximately 6.3 ms and memory usage of 13.2 MB. The proposed MFE-based approach improves accuracy by +2.0% (from 95.0% to 97.0%), reduces total inference time by 89.0% (from 37.4 ms to 4.1 ms), and decreases memory usage by 86.4% (from 14 MB to 1.9 MB). Hyperparameter tuning results demonstrate performance improvements of 1.3% for the baseline model and 1.0% for the proposed model.

Keywords: Audio Classification, Emergency Vehicles, Mel-Filterbank Energy, Ensemble Learning, 1D CNN