

## INTISARI

### **Deteksi Audio *Deepfake* Menggunakan *Constant Q Transform* dan Teknik Augmentasi Data Dinamis**

Oleh

Fayza Nizma Safaya Harda  
21/474396/PA/20483

Dengan perkembangan teknologi kecerdasan buatan (AI), teknologi deepfake audio seperti *Text-to-Speech* (TTS) dan *Voice Conversion* (VC) telah digunakan untuk meningkatkan kenyamanan dalam berbagai aplikasi. Namun, teknologi ini memiliki potensi risiko yang besar, termasuk penyalahgunaan untuk kejahatan seperti penipuan finansial. Salah satu tantangan utama dalam deteksi audio *deepfake* adalah kemampuan model untuk mengenali berbagai metode pemalsuan yang belum diketahui sebelumnya, terutama ketika model diuji dengan data dari domain atau jenis serangan yang berbeda.

Penelitian ini mengusulkan kombinasi metode ekstraksi fitur menggunakan *Constant Q Transform* (CQT), klasifikasi dengan *Light Convolutional Neural Network* (LCNN), dan teknik augmentasi data seperti *MixUp*, *CutMix*, dan *ShuffleMix*. Pendekatan ini dirancang untuk meningkatkan kemampuan generalisasi model deteksi audio deepfake dengan menghasilkan distribusi data yang lebih beragam serta mengurangi risiko *overfitting*. Model diuji menggunakan dataset ASVspoof 2019 *Logical Access*, sedangkan dataset ASVspoof 2021 *Deepfake* digunakan untuk pengujian *cross-dataset*.

Hasil penelitian menunjukkan bahwa kombinasi metode yang diusulkan berhasil meningkatkan performa deteksi audio *deepfake*. Teknik augmentasi data terbukti efektif, dengan *CutMix* memberikan penurunan *Equal Error Rate* (EER)  $\pm 1\%$  pada dataset ASVspoof 2019 dan *ShuffleMix* memberikan penurunan EER  $\pm 4\%$  pada dataset ASVspoof 2021.

**Kata Kunci:** *audio deepfake detection, constant q transform, light convolutional neural network (lcnn), data augmentation, voice conversion*

## ABSTRACT

### **Audio Deepfake Detection Using Constant Q Transform and Dynamic Data Augmentation Techniques**

by

Fayza Nizma Safaya Harda  
21/474396/PA/20483

With the advancement of Artificial Intelligence (AI) technology, deepfake audio technologies such as Text-to-Speech (TTS) and Voice Conversion (VC) have been utilized to enhance convenience in various applications. However, these technologies pose significant risks, including misuse for crimes such as financial fraud. One of the primary challenges in detecting audio deepfakes is the ability of models to recognize various forgery methods that have not been encountered before, especially when tested on data from different domains or types of attacks.

This study proposes a combination of methods involving feature extraction using Constant Q Transform (CQT), classification with Light Convolutional Neural Network (LCNN), and data augmentation techniques such as MixUp, CutMix, and ShuffleMix. This approach is designed to improve the generalization capability of audio deepfake detection models by generating a more diverse data distribution and reducing the risk of overfitting. The model was tested using the ASVspoof 2019 Logical Access dataset, while the ASVspoof 2021 Deepfake dataset was used for cross-dataset evaluation.

The results demonstrate that the proposed methods successfully enhanced the performance of audio deepfake detection. Data augmentation techniques proved effective, with CutMix reducing the Equal Error Rate (EER) by  $\pm 1\%$  on the ASVspoof 2019 dataset and ShuffleMix reducing the EER by  $\pm 4\%$  on the ASVspoof 2021 dataset.

**Keywords:** *audio deepfake detection, constant q transform, light convolutional neural network (lcnn), data augmentation, voice conversion*