



INFORMATICS  
INSTITUTE OF  
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**Robust deepfake audio detection method  
with tackling the real-world background noises**

A Dissertation by

Ms. S.A. Nanduni Kaveesha

Supervised by

Mr. Nipuna Senanayake

Submitted in partial fulfillment of the requirements for the BSc (Hons) Computer Science degree at the University of Westminster.

**April 2023**

## Abstract

There is fast development in the artificial intelligence field nowadays, and with these rapid developments, deep fake audio, images, and videos are becoming difficult to recognize by basic human intelligence. To solve this issue with the use of recently developed technologies, there are deep-fake detection systems to identify these fake images, videos, and audio. But there is a significant deficiency in proper deepfake audio detection methods in the deepfake audio field. There are only a few research papers and systems for deep-fake audio detection methods that can also tackle the background noises of the audio file. Fulfilling this research gap by developing a deep-fake audio detection system that also tackles background noises is the main purpose of this project.

To achieve this goal, gathering suitable audio datasets was the first step. The author has found a standard audio dataset that consists of 10,000 real and fake audio files. In addition to that, a background noises audio dataset, which consists of dog, bird, and rain noises, is also collected to train the model to tackle the background noises. After that, the main audio dataset is created by mixing those two audio datasets and having it approved by the supervisor. For the audio pre-processing part, suitable and standard methods have been used, and finally, the audio files convert to images, which then convert to numpy arrays for feature extraction purposes. Before building this model, the author experimented with autoencoders, spectral subtraction, CNN, and RNN, and finally chose the ensemble model technique for the system to build a better-performing model with good accuracy. And for the base model, pre-trained MobileNetV2 is chosen considering it is a fast, effective, and lightweight model with a good reputation for classification purposes. And MobileNetV2 model also has been trained with many classification datasets before. In addition to that base model, some other layers like dense layers, global average pooling layers, batch normalization layers, and dropout layers have been added accordingly.

This proposed, designed, and developed deepfake audio detection system is able to accurately and efficiently detect deepfake audio files while tackling the background real-world noises successfully. And the generalization and robustness of the model have also been considered and improved, as expected. With that information, it is undeniable that this developed deep-fake audio detection system is a fair contribution to the recognized research gap.

**CNN – Convolutional Neural Network**

**RNN – Recurrent Neural Network**

**Spectral subtraction – a method for reducing background noise by estimating the spectral characteristics of the noise and then subtracting it from the audio signal.**

**Autoencoder/decoder – is an artificial neural network that are widely used for anomaly detection, dimensionality reduction and data denoising.**