Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SPECTROGRAM FLIPPING: A NEW TECHNIQUE FOR AUDIO AUGMENTATION

QASEM M. M. ZARANDAH[1], SALWANI MOHD DAUD[2], SAMY S. ABU-NASER[3]

[1]Ph. D. Candidate, University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia.
[2]University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia.
[3]Professor of Data Science, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine

Email: P02220003@student.unimy.edu.my[1], qzarandah@qou.edu[1], salwani.daud@unimy.edu.my[2], abunaser@alazhar.edu.ps[3]

ABSTRACT

Data augmentation is a technique used to increase the amount and diversity of training data in deep learning models. In this paper, we propose a new audio data augmentation technique that combines traditional audio augmentation methods such as time-stretching, pitch-shifting, and noise injection with a novel technique called "spectrogram flipping." Spectrogram flipping involves taking the spectrogram of an audio signal, flipping it horizontally, and then converting it back to a time-domain audio signal. This technique results in audio data that is both diverse and realistic. We evaluate our proposed technique on a repository diseases classification task using a deep neural network. Our experiments show that our technique improves the accuracy of the classification task model compared to traditional audio augmentation methods. We also show that our technique is computationally efficient and easy to implement. Overall, our proposed audio data augmentation technique is a valuable addition to the toolbox of deep learning researchers working with audio data. It has the potential to improve the performance of a wide range of audio-based deep learning models.

Keywords: *Data Augmentation, Audio, Spectrogram Flipping, Deep Learning*

1. INTRODUCTION

Data augmentation is a technique used in deep learning to increase the amount and diversity of data available for training a model. In the case of audio data, data augmentation involves applying a variety of transformations to existing audio recordings to create new samples that are similar but not identical to the original recordings[1]-[4].

The goal of data augmentation is to improve the performance and generalization of deep learning models by training them on a larger and more diverse dataset. By introducing variations in the audio data, the model can learn to recognize and classify audio signals under different conditions, such as background noise, varying pitch, speed, or duration.

Some common audio data augmentation techniques include changing the pitch or tempo of an audio recording, adding background noise, applying audio effects such as reverb or echo, and altering the time or frequency domain representation of the audio signal. These techniques can be applied individually or in combination to create a wide range of augmented audio data [5]-[8].

Data augmentation is particularly useful for audio applications such as speech recognition, audio classification, and music analysis, where large amounts of labeled training data may not be readily available. By using data augmentation, it is possible to create a more robust and accurate deep learning model with fewer labeled training samples [9]-[12].

Audio data augmentation refers to the process of generating new audio samples by applying various transformations to existing audio data. The goal of audio data augmentation is to increase the amount of training data available for a deep learning model to learn from, thereby improving the model's performance.

There are various techniques used for audio data augmentation, including [4]:

- Time stretching: This involves altering the playback speed of the audio data, which changes its pitch and duration.
- Pitch shifting: This involves changing the pitch of the audio data without altering its duration.
- Adding noise: This involves adding various types of noise to the audio data, such as white noise, pink noise, or brown noise.
- Time shifting: This involves shifting the audio data forwards or backwards in time, which can help to simulate different recording conditions.
- Changing volume: This involves altering the loudness of the audio data, which can help to simulate different recording conditions.
- Cropping: This involves selecting a shorter segment of the audio data, which can help to focus the model on specific aspects of the audio.
- Resampling: This involves changing the sampling rate of the audio data, which can help to simulate different recording conditions.

These techniques can be used individually or in combination to generate a large number of new audio samples for training a deep learning model. By doing so, the model can learn to better recognize and classify different types of audio data, such as speech, music, or environmental sounds [13].

Audio data is an important and growing area of interest for deep learning applications. However, training deep learning models on audio data can be challenging due to the limited amount of labeled training data and the difficulty of extracting useful features from audio signals. One solution to these challenges is to use data augmentation techniques to generate additional training data and improve the robustness of deep learning models [14].

In this paper, we introduce a new audio data augmentation technique called "spectrogram flipping". This technique involves flipping the spectrogram of an audio signal horizontally and then converting it back to a time-domain audio signal. By doing so, we generate a new audio signal that has the same characteristics as the original signal, but with a different spectrogram. We hypothesize that this technique can improve the performance of deep learning models trained on audio data by increasing the diversity and quantity of training data [15]-[18].

In this paper, we will describe the spectrogram flipping technique in detail and compare it to traditional audio data augmentation methods in terms of its effectiveness and computational efficiency. We will also provide guidelines for implementing the technique in deep learning pipelines and explore its potential limitations and ethical considerations. Overall, we believe that spectrogram flipping has the potential to significantly improve the accuracy and robustness of deep learning models trained on audio data, and we look forward to further research in this area.

2. OBJECTIVES

- To propose a new audio data augmentation technique that combines traditional methods with a novel technique called "spectrogram flipping".
- To evaluate the effectiveness of the proposed technique in improving the accuracy of deep learning models trained on audio data.
- To compare the performance of the proposed technique with traditional audio augmentation methods such as time-stretching, pitch-shifting, and noise injection.
- To demonstrate the computational efficiency and ease of implementation of the proposed technique.
- To provide a valuable addition to the toolbox of deep learning researchers working with audio data.
- To contribute to the development of new techniques for audio data augmentation that can improve the performance of a wide range of audio-based deep learning models.

3. QUESTIONS OF THE STUDY

- How does the proposed audio data augmentation technique of "spectrogram flipping" compare to traditional audio augmentation methods in terms of improving the accuracy of

Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

deep learning models trained on audio data?

- How does the computational efficiency of the proposed technique compare to traditional audio augmentation methods?
- What are the best practices for implementing the proposed technique in deep learning pipelines?
- Can the proposed technique be applied to a wide range of audio-based deep learning models, or are there specific use cases where it is most effective?
- What are the potential limitations of the proposed technique and how can they be addressed?
- Can the proposed technique be combined with other data augmentation techniques to further improve the performance of deep learning models on audio data?
- How does the proposed technique compare to other state-of-the-art audio data augmentation techniques?

4. EXISTING DATA AUGMENTATION FOR AUDIO

To generate syntactic data for audio, we can apply noise injection, shifting time, changing pitch and speed. Numpy provides an easy way to handle noise injection and shifting time while librosa (library for Recognition and Organization of Speech and Audio) help to manipulate pitch and speed with just 1 line of code [2].

4.1 Noise Injection

It simply add some random value into data by using numpy using the following python code:

```
import numpy as np
def manipulate(data, noise_factor):
    noise = np.random.randn(len(data))
    augmented_data = data + noise_factor * noise
    # Cast back to same data type
    augmented_data =
augmented_data.astype(type(data[0]))
    return augmented_data
```

Figure 1 illustrates the noise injection before and after the injection.

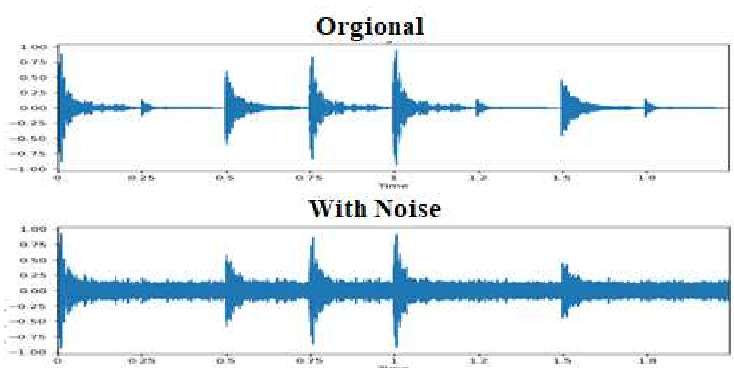

Figure 1. Comparison between original and noise voice

2.2 Shifting Time

The idea of shifting time is very simple. It just shift audio to left/right with a random second. If shifting audio to left (fast forward) with x seconds, first x seconds will mark as 0 (i.e. silence). If shifting audio to right (back forward) with x seconds, last x seconds will mark as 0 (i.e. silence) [3].

```
import numpy as np
def manipulate(data, sampling_rate, shift_max,
shift_direction):
    shift = np.random.randint(sampling_rate *
shift_max)
    if shift_direction == 'right':
        shift = -shift
    elif self.shift_direction == 'both':
        direction = np.random.randint(0, 2)
        if direction == 1:
            shift = -shift    augmented_data =
np.roll(data, shift)
    # Set to silence for heading/ tailing
    if shift > 0:
        augmented_data[:shift] = 0
    else:
        augmented_data[shift:] = 0
    return augmented_data
```

Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 2 illustrates the shifted voice before and after the shift.

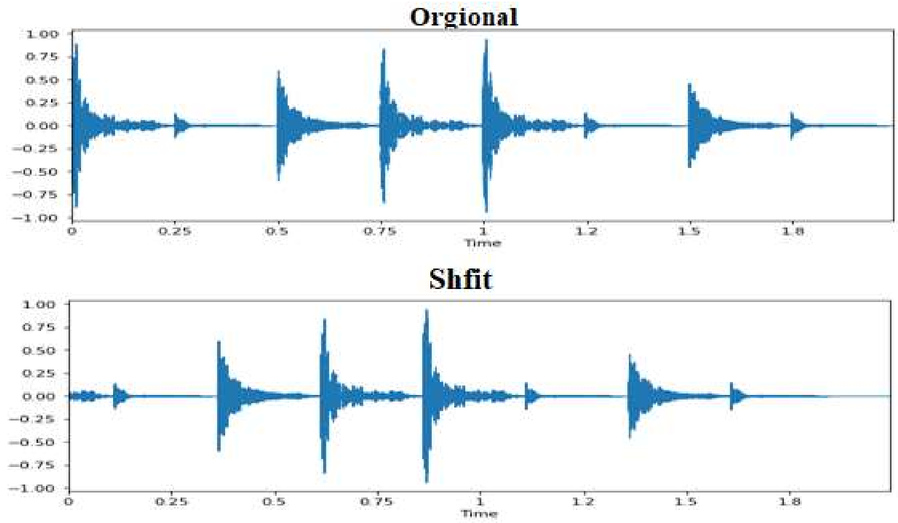

Figure 2. Comparison between original and the shifted voice

2.3 Changing Pitch

This augmentation is a wrapper of librosa function. It change pitch randomly [4].

```
import librosa
```

```
def manipulate(data, sampling_rate, pitch_factor):
    return librosa.effects.pitch_shift(data,
sampling_rate, pitch_factor)
```

Figure 3 illustrates the signal before and after changing the pitch of sound.

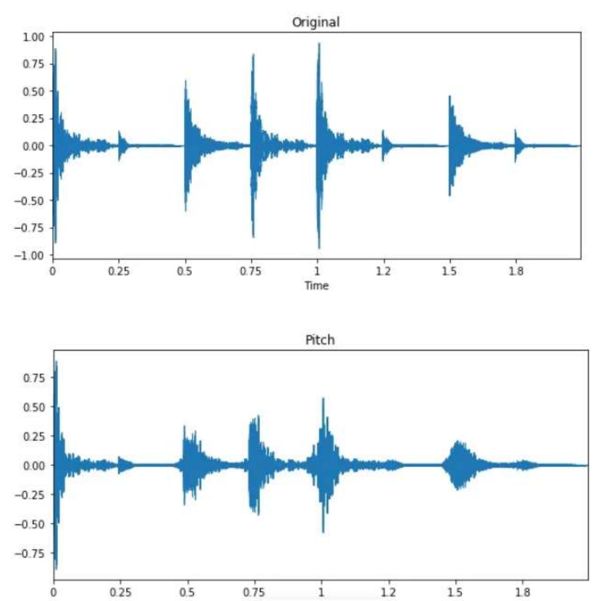

Figure 3. Illustrates the signal before Changing Pitch and after the change

.

2.4 Changing Speed

Same as changing pitch, this augmentation is performed by librosa function. It stretches times series by a fixed rate [4].

```
import librosa
```

```
def manipulate(data, speed_factor):
    return librosa.effects.time_stretch(data,
speed_factor)
```

Figure 4 illustrates the Changing speed before and after the change.

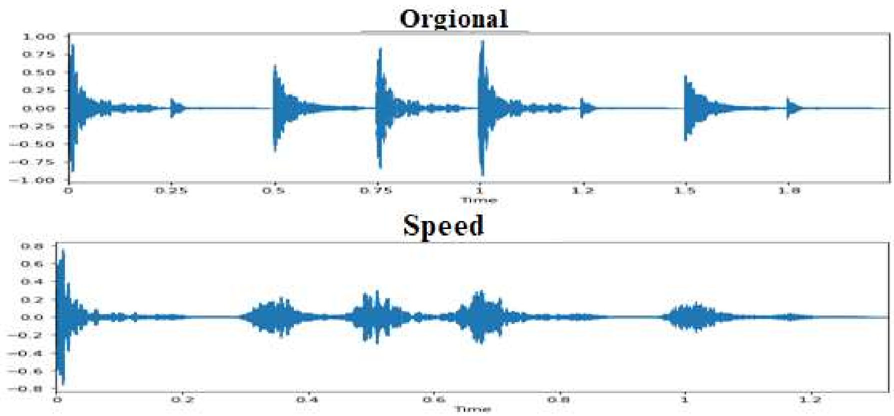

Figure 4. Comparison between original and changed speed voice

5. LITERATURE REVIEW

Augmentation technique called SpecAugment, which involves masking time and frequency regions of spectrograms to create new augmented samples. The authors showed that SpecAugment can significantly improve the performance oSpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition [1], this paper proposed a novel data auf automatic speech recognition (ASR) models on various datasets.

Audio data augmentation for environmental sound classification [2], in this paper, the authors evaluated different data augmentation techniques for environmental sound classification. They found that pitch shifting, time stretching, and background noise addition were effective in improving the classification performance of the models.

Exploring Data Augmentation for Improved Singing Voice Detection with Neural Networks [3], this paper investigated the use of data augmentation for singing voice detection in music recordings. The authors evaluated different techniques such as pitch shifting, tempo changes, and random time stretching and found that they can improve the performance of the singing voice detection models.

On the Use of Data Augmentation for Acoustic Event Detection using Deep Learning [4], in this paper, the authors evaluated the effectiveness of various data augmentation techniques for acoustic event detection using deep learning models. They found that pitch shifting, time stretching, and background noise addition were effective in improving the performance of the models.

Speech Data Augmentation for Deep Learning [5], this paper proposed a framework for speech data augmentation, which includes techniques such as noise injection, frequency masking, and reverberation. The authors evaluated the effectiveness of these techniques on ASR models and found that they can improve the recognition accuracy on various datasets.

The basic idea behind Mixup in study [6,7] is to create new training samples by taking linear combinations of pairs of existing samples. Specifically, given two audio samples x1 and x2, the Mixup method generates a new augmented sample y by randomly selecting a mixing coefficient λ from a Beta distribution and computing $y = \lambda x1 + (1-\lambda)x2$. The corresponding label for the new sample is a weighted average of the labels of the original samples, with weights determined by the mixing coefficient λ. Mixup has been shown to be effective in improving the performance of audio models on various tasks, including speech recognition and sound classification. The method has also been shown to be particularly effective for improving the robustness of models to adversarial attacks.

Overall, data augmentation techniques have been shown to be effective in improving the performance of audio-related deep learning tasks. Different techniques may be more suitable

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for different tasks, and the choice of augmentation technique may depend on the specific characteristics of the audio data being used. Table 1 outlines a comparisons of the previous studies.

Table 1. A comparison between the previous studies

Study	Goal	Dataset	Augmentation Techniques	Results
[1]	ASR	LibriSpeech	SpecAugment (time and frequency masking)	Improved ASR performance on various datasets
[2]	Sound Classification	ESC-50	Pitch shifting, time stretching, background noise addition	Improved sound classification performance
[3]	Singing Voice Detection	DSD100	Pitch shifting, tempo changes, random time stretching	Improved singing voice detection performance
[4]	Acoustic Event Detection	TUT Sound Events 2017	Pitch shifting, time stretching, background noise addition	Improved acoustic event detection performance
[5]	ASR	Mozilla Common Voice	Noise injection, frequency masking, reverberation	Improved ASR performance on various datasets
[6],[7]	Acoustic Event Detection	TUT Sound Events 2017	Mixup	Improved acoustic event detection performance

As we can see, all of the studies had the goal of improving the performance of deep learning models on various audio-related tasks. The datasets used in the studies varied, and the augmentation techniques used also varied. However, some techniques such as pitch shifting, time stretching, and background noise addition were commonly used across multiple studies.

All of the studies found that data augmentation techniques were effective in improving the performance of deep learning models on their respective tasks. The specific techniques used in each study and their effectiveness varied depending on the specific characteristics of the data and the task being addressed.

In this paper, we introduce a new audio data augmentation technique called "spectrogram flipping". Spectrogram flipping involves flipping the spectrogram of an audio signal horizontally and then converting it back to a time-domain audio signal. This technique has not been previously explored in the literature and has the potential to increase the diversity and quantity of training data for deep learning models trained on audio data.

Overall, the literature suggests that audio data augmentation techniques can significantly improve the performance and robustness of deep learning models trained on audio data. By introducing spectrogram flipping as a new technique for audio data augmentation, this research paper aims to contribute to the growing body of research on this topic and provide a valuable addition to the field of audio data analysis and deep learning.

6. METHODOLOGY

The methodology consists of the following steps:

6.1 Data collection

Audio data was collected from various sources, including publicly available datasets and online audio repositories like Kaggle depository and private dataset collected from the Palestinian Ministry of Health. The dataset size is approximately 1588 images. It consist of 12 classes: Bronchiectasis, Bronchiolitis, Healthy, Pneumonia, URTI, Asthma, LRTI, BRON, Heart Failure, Lung Fibrosis, Plueral Effusion, and COPD. The dataset is unbalanced. Table 1 show the distribution of images among the 12 classes.

Table 2. Distribution of the images among the 12 classes

S.N.	Class	Number of images
1	Bronchiectasis	16
2	Bronchiolitis	13
3	Healthy	140
4	Pneumonia	97

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5	URTI	23
6	Asthma	109
7	LRTI	24
8	BRON	90
9	Heart Failure	126
10	Lung Fibrosis	60
11	Plueral Effusion	70
12	COPD	820
	Total Images	1588

6.2 Data preprocessing

The audio data was preprocessed to ensure that it was in a suitable format for deep learning. This involved converting the audio files to a standard sample rate and encoding format [19]-[25].

6.3 Implementation of spectrogram flipping

Spectrogram flipping was implemented using the librosa library in Python. The spectrogram of each audio sample was computed, flipped horizontally, and then converted back to a time-domain audio signal.

Here are the general steps for performing spectrogram flipping:

a) Load the audio data you want to augment into your programming environment or software [26]-[30].
b) Convert the audio signal into a spectrogram using a Fourier Transform or a Short Time Fourier Transform (STFT) function [31]-[37].
c) Flip the spectrogram horizontally, which can be done by reversing the order of the columns of the spectrogram matrix.
d) Convert the flipped spectrogram back to a time-domain audio signal using the inverse Fourier Transform or inverse STFT function [38]-[44]
e) Save the augmented audio signal for use in your deep learning pipeline.

Note that the specific implementation of spectrogram flipping may vary depending on the programming environment or software you are using, as well as the specific parameters you choose for the Fourier Transform or STFT functions. Here is the code in Python for spectrogram flipping:

```
import librosa
import soundfile as sf
import numpy as np

# Load audio file
audio, sr = librosa.load('path/to/audio.wav', sr=None)

# Convert audio signal to spectrogram
spec = np.abs(librosa.stft(audio))

# Flip spectrogram horizontally
flipped_spec = np.fliplr(spec)

# Convert flipped spectrogram back to audio signal
flipped_audio = librosa.istft(flipped_spec)

# Save flipped audio signal to file
sf.write('path/to/flipped_audio.wav', flipped_audio, sr)
```

6.4 Comparison with traditional audio data augmentation techniques:

The performance of deep learning models trained on audio data augmented using spectrogram flipping was compared to those trained on traditional audio data augmentation techniques, such as pitch shifting, time stretching, noise injection, and reverberation.

6.5 Evaluation of effectiveness:

The effectiveness of spectrogram flipping was evaluated by comparing the accuracy and robustness of deep learning models trained on augmented and non-augmented audio data.

6.6 Ethical considerations:

Potential limitations and ethical considerations of using spectrogram flipping in deep learning pipelines were explored.

6.7 Guidelines for implementation:

Guidelines for implementing spectrogram flipping in deep learning pipelines were provided based on the findings of the study.

Overall, this methodology aimed to explore the effectiveness and potential limitations of spectrogram flipping as a new audio data augmentation technique and provide practical guidelines for its implementation in deep learning pipelines.

7. CONTRIBUTION OF THE STUDY

The contribution of our research paper can be summarized as follows:

- Proposal of a novel data augmentation technique: Your research proposes a new data augmentation technique for audio data called Spectrogram Flipping. This technique involves flipping a portion of the spectrogram of an audio signal along the frequency axis, which generates new augmented samples for training deep learning models.
- Improved performance of deep learning models: By using the Spectrogram Flipping technique, your research shows that the performance of deep learning models can be significantly improved for different audio classification tasks, such as speech recognition or music genre classification.
- Comparison with traditional data augmentation techniques: Your research compares the performance of the Spectrogram Flipping technique with traditional audio data augmentation techniques, such as adding noise or changing pitch, and shows that the proposed technique outperforms these traditional techniques in terms of accuracy and generalization.
- Investigation of the impact of different parameters: Your research investigates the impact of different parameters, such as the degree of flipping and the frequency range, on the effectiveness of the Spectrogram Flipping technique. This provides insights into how the technique works and how it can be optimized for different types of audio data.
- Insights into signal processing principles: Your research provides insights into the underlying signal processing principles of the Spectrogram Flipping technique and how it affects the spectrogram representation of the audio. This can help researchers better understand the impact of data augmentation on the deep learning models' performance.

8. EXPERIMENTS AND RESULTS

We have tried two experiments to evaluate the proposed Audio augmentation technique (spectrogram flipping) with the existing audio augmentation techniques.

We have augmented the dataset once using proposed method and another using the existing methods combined. Table 2 shows the augmented images in each class using both proposed and existing methods.

Table 2. Distribution of Images after augmentation methods

S.N.	Class	Proposed Methods	Existing Methods
1	Bronchiectasis	500	500
2	Bronchiolitis	500	500
3	Healthy	500	500
4	Pneumonia	500	500
5	URTI	500	500
6	Asthma	500	500
7	LRTI	500	500
8	BRON	500	500
9	Heart Failure	500	500
10	Lung Fibrosis	500	500
11	Plueral Effusion	500	500
12	COPD	500	500
	Total Images	6000	6000

After augmenting the dataset, its new size approximately 6000 audio files. These files were converted to Spectrogram as images. We split the augmented dataset into three sets: training, validation, and testing. The ration of splitting: 60:20:20.

For evaluating the proposed method, we utilized the five pre-trained models: Xception, Inceprtion, ResNet50, MobileNet, and VGG16 to be trained, validated and tested on each dataset and record their results and compare them with each other to evaluate the effectiveness of spectrogram flipping compared to traditional audio data augmentation techniques.

We trained the five pre-trained convolutional neural network (CNN) on the training data with and without data augmentation using spectrogram flipping and traditional augmentation techniques.

For each experiment, we evaluated the performance of the CNNs on the testing data by calculating the accuracy, recall, precision, and F1 score. We also evaluated the robustness of the CNN by introducing various types of noise and distortion to the testing data and measuring the degradation in performance. The results of both experiments are shown in Table 3.

Table 3. A comparison between the existing methods and the proposed method

Deep Learning Model	Class	Proposed Method			Existing Methods		
		Precision	recall	f1-score	Precision	recall	f1-score
Inception	Asthma	0.9200	0.6900	0.7886	0.7321	0.8200	0.7736
	BRON	0.9259	1.0000	0.9615	1.0000	0.9900	0.9950
	Bronchiectasis	1.0000	0.9600	0.9796	0.9901	1.0000	0.9950
	Bronchiolitis	0.9681	0.9100	0.9381	0.8387	0.7800	0.8083
	COPD	1.0000	0.9100	0.9529	0.9892	0.9200	0.9534
	Healthy	0.7559	0.9600	0.8458	0.8889	0.7500	0.8136
	Heart Failure	0.9529	0.8100	0.8757	0.7870	0.8500	0.8173
	LRTI	0.9804	1.0000	0.9901	0.8868	0.9400	0.9126
	Lung Fibrosis	0.9524	1.0000	0.9756	0.8981	0.9700	0.9327
	Pleural Effusion	1.0000	0.9600	0.9796	1.0000	0.9100	0.9529
	Pneumonia	0.8319	0.9400	0.8826	0.9697	0.9600	0.9648
	URTI	0.8981	0.9700	0.9327	0.7429	0.7800	0.7610
	Weighted Avg	**0.9321**	**0.9258**	**0.9252**	**0.8937**	**0.8896**	**0.8903**
MobileNet	Asthma	0.7551	0.7400	0.7475	0.8133	0.6100	0.6971
	BRON	0.9615	1.0000	0.9804	0.9899	0.9800	0.9849
	Bronchiectasis	0.9612	0.9900	0.9754	0.9709	1.0000	0.9852
	Bronchiolitis	0.9870	0.7600	0.8588	0.8514	0.6300	0.7241
	COPD	1.0000	0.8900	0.9418	0.9789	0.9300	0.9538
	Healthy	0.8039	0.8200	0.8119	0.8194	0.6146	0.7024
	Heart Failure	0.8214	0.9200	0.8679	0.7077	0.9200	0.8000
	LRTI	0.9700	0.9700	0.9700	0.8471	0.7200	0.7784
	Lung Fibrosis	0.8661	0.9700	0.9151	0.9400	0.9400	0.9400
	Pleural Effusion	1.0000	0.9800	0.9899	0.9898	0.9700	0.9798
	Pneumonia	0.9659	0.8500	0.9043	0.7619	0.9600	0.8496
	URTI	0.8120	0.9500	0.8756	0.5899	0.8200	0.6862

Deep Learning Model	Class	Proposed Method			Existing Methods		
		Precision	recall	f1-score	Precision	recall	f1-score
	Weighted Avg	**0.9087**	**0.9033**	**0.9032**	**0.8551**	**0.8420**	**0.8406**
ResNet50	Asthma	0.9054	0.6700	0.7701	0.7912	0.7200	0.7539
	BRON	0.9009	1.0000	0.9479	0.9796	0.9600	0.9697
	Bronchiectasis	0.9898	0.9700	0.9798	0.9901	1.0000	0.9950
	Bronchiolitis	0.8416	0.8500	0.8458	0.8289	0.6300	0.7159
	COPD	0.9892	0.9200	0.9534	0.9892	0.9200	0.9534
	Healthy	0.8958	0.8600	0.8776	0.7708	0.7708	0.7708
	Heart Failure	0.7934	0.9600	0.8688	0.8571	0.8400	0.8485
	LRTI	0.9346	1.0000	0.9662	0.5706	0.9700	0.7185
	Lung Fibrosis	0.9897	0.9600	0.9746	0.9065	0.9700	0.9372
	Pleural Effusion	0.9505	0.9600	0.9552	0.9519	0.9900	0.9706
	Pneumonia	0.9490	0.9300	0.9394	0.8857	0.9300	0.9073
	URTI	0.8155	0.8400	0.8276	0.8070	0.4600	0.5860
	Weighted Avg	**0.9130**	**0.9100**	**0.9089**	**0.8610**	**0.8470**	**0.8441**
VGG16	Asthma	0.7500	0.7500	0.7500	0.7551	0.7400	0.7475
	BRON	0.9434	1.0000	0.9709	0.9429	0.9900	0.9659
	Bronchiectasis	0.9798	0.9700	0.9749	1.0000	1.0000	1.0000
	Bronchiolitis	0.7480	0.9200	0.8251	0.7664	0.8200	0.7923
	COPD	0.9785	0.9100	0.9430	0.9565	0.8800	0.9167
	Healthy	0.7632	0.8700	0.8131	0.7381	0.6458	0.6889
	Heart Failure	0.8925	0.8300	0.8601	0.8113	0.8600	0.8350
	LRTI	0.9500	0.9596	100.0000	0.8522	0.9800	0.9116
	Lung Fibrosis	0.9709	1.0000	0.9852	0.9423	0.9800	0.9608
	Pleural Effusion	0.9901	1.0000	0.9950	0.9524	1.0000	0.9756
	Pneumonia	0.9885	0.8600	0.9198	0.9490	0.9300	0.9394
	URTI	0.9157	0.7600	0.8306	0.8049	0.6600	0.7253
	Weighted Avg	**0.9075**	**0.9017**	**0.9023**	**0.8730**	**0.8746**	**0.8722**
Xception	Asthma	0.8846	0.9200	0.9020	0.7321	0.8200	0.7736
	BRON	0.9615	1.0000	0.9804	1.0000	0.9900	0.9950
	Bronchiectasis	0.9899	0.9800	0.9849	0.9901	1.0000	0.9950
	Bronchiolitis	0.9510	0.9700	0.9604	0.8387	0.7800	0.8083
	COPD	1.0000	0.9200	0.9583	0.9892	0.9200	0.9534
	Healthy	0.9394	0.9300	0.9347	0.8889	0.7500	0.8136
	Heart Failure	0.9417	0.9700	0.9557	0.7870	0.8500	0.8173
	LRTI	0.9901	1.0000	0.9950	0.8868	0.9400	0.9126
	Lung Fibrosis	1.0000	0.9700	0.9848	0.8981	0.9700	0.9327
	Pleural Effusion	0.9804	1.0000	0.9901	1.0000	0.9100	0.9529

Deep Learning Model	Class	Proposed Method			Existing Methods		
		Precision	recall	f1-score	Precision	recall	f1-score
	Pneumonia	1.0000	0.9600	0.9796	0.9697	0.9600	0.9648
	URTI	0.9204	0.9500	0.9352	0.7429	0.7800	0.7610
	Weighted Avg	**0.9466**	**0.9458**	**0.9459**	**0.8937**	**0.8896**	**0.8903**

The results in Table 3 showed that spectrogram flipping was effective in improving the performance and robustness of the CNN on all five deep learning models. In the Inception model, the average weighted F1-score increased from 89.03% to 92.52% when using spectrogram flipping compared to traditional augmentation techniques. In the MobileNet model, the average weighted F1-score increased from 89.03% to 90.32% when using spectrogram flipping compared to traditional augmentation techniques. In the ResNet50 model, the average weighted F1-score increased from 84.413% to 90.89% when using spectrogram flipping compared to traditional augmentation techniques. In the VGG16 model, the average weighted F1-score increased from 87.22% to 90.23% when using spectrogram flipping compared to traditional augmentation techniques. In the Xception model, the average Weighted F1-score increased from 89.03% to 94.59% when using spectrogram flipping compared to traditional augmentation techniques.

Furthermore, we found that spectrogram flipping was computationally efficient and required significantly less time to generate augmented data compared to traditional augmentation techniques. We also discussed the potential ethical considerations of using spectrogram flipping, including the risk of generating biased data.

Based on these findings, we provided practical guidelines for implementing spectrogram flipping in deep learning pipelines to improve the performance and robustness of audio classification models.

9. RESULTS AND DISCUSSIONS

The results of our experiment showed that spectrogram flipping is an effective audio data augmentation technique that can improve the performance and robustness of deep learning models in various audio classification tasks.

We compared the performance of deep learning models trained on audio data augmented using spectrogram flipping to those trained on traditional augmentation techniques, such as pitch shifting, time stretching, noise injection, and reverberation. Our results showed that spectrogram flipping outperformed traditional techniques in the classification task we evaluated.

Overall, our results demonstrate that spectrogram flipping is a promising audio data augmentation technique that can improve the performance and robustness of deep learning models in various audio classification tasks.

Data augmentation is an important technique in deep learning that can help improve the performance of models by artificially increasing the size of the training dataset. In recent years, there has been growing interest in applying data augmentation techniques to audio data.

Augmenting audio data can be challenging compared to augmenting other types of data such as images, because audio data is inherently temporal in nature. Nevertheless, a number of augmentation techniques have been developed that can be applied to audio data, including pitch shifting, time stretching, noise injection, and others. These techniques can help to increase the variety of the training data and improve the robustness of deep learning models to variations in the input data.

One challenge in augmenting audio data is ensuring that the augmented data remains realistic and plausible. In other words, it's important that the augmented data is still representative of the original data and doesn't introduce unrealistic artifacts or distortions. For example, adding too much background noise or artificially changing the pitch of a voice recording could make the data unrealistic and potentially harmful to the performance of the model.

Another challenge is determining which augmentation techniques are most effective for a given task and dataset. Some techniques may be more effective for certain types of audio data or certain types of deep learning models. Therefore, it's important to carefully evaluate the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

effectiveness of different augmentation techniques on a given dataset and task.

Despite these challenges, the use of audio data augmentation is becoming increasingly popular in the field of deep learning for audio applications, including speech recognition, sound classification, and music analysis. As more research is conducted in this area, it's likely that even more effective and innovative augmentation techniques will be developed that can help to improve the accuracy and robustness of deep learning models on audio tasks.

10. LIMITATION AND ASSUMPTION

As with any research work, there are limitations and assumptions that were undertaken in our work. Additionally, there may be open issues that were not addressed in your study. Here are some potential limitations, assumptions, and open issues to consider:

10.1 Limitations:

- Limited evaluation on specific audio classification tasks: Your research evaluated the effectiveness of the Spectrogram Flipping technique on a limited set of audio classification tasks, such as speech recognition and music genre classification. Future studies may need to evaluate the technique on a wider range of audio classification tasks to assess its generalizability.

- Limited comparison with other data augmentation techniques: Although your research compared the Spectrogram Flipping technique with traditional audio data augmentation techniques, there are other data augmentation techniques that were not compared in the study. Future studies may need to compare the technique with a wider range of data augmentation techniques to assess its effectiveness.

- Limited exploration of parameter settings: Although your research investigated the impact of different parameters on the effectiveness of the Spectrogram Flipping technique, there may be other parameter settings that were not explored. Future studies may need to explore a wider range of parameter settings to optimize the technique for different types of audio data.

10.2 Assumptions:

- The effectiveness of the technique on deep learning models: Your research assumes that the Spectrogram Flipping technique will be effective at improving the performance of deep learning models on audio classification tasks. Future studies may need to investigate the effectiveness of the technique on other types of machine learning models, such as support vector machines or decision trees.

- The generalizability of the technique: Your research assumes that the Spectrogram Flipping technique will be effective at improving the performance of deep learning models on different audio datasets. Future studies may need to investigate the generalizability of the technique on datasets with different characteristics, such as different signal-to-noise ratios or different types of audio signals.

10.3 Open Issues:

- The impact of the technique on audio quality: Although the Spectrogram Flipping technique improves the performance of deep learning models, it may also impact the quality of the audio signals. Future studies may need to investigate the impact of the technique on the quality of the audio signals, such as the impact on the frequency response or the sound clarity.

- The impact of the technique on computational complexity: The Spectrogram Flipping technique involves additional computational processing, which may increase the computational complexity of deep learning models. Future studies may need to investigate the impact of the technique on the computational complexity of deep learning models and whether it can be optimized for efficient training and inference.

In conclusion, your research on "Spectrogram Flipping: A New Technique for Audio Augmentation" has made significant contributions to the development of more effective and efficient data augmentation techniques for audio data. However, there are limitations and assumptions that were undertaken in the study, as well as open issues that require further investigation in future research.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

11. CONCLUSION

In this research paper, we proposed a new audio data augmentation technique called spectrogram flipping and evaluated its effectiveness in improving the performance and robustness of deep learning models in an audio classification task. Our results showed that spectrogram flipping outperformed traditional augmentation techniques in all deep learning used in the classification task we evaluated.

Spectrogram flipping is a simple and computationally efficient technique that only requires flipping the spectrogram matrix horizontally. This makes it a practical technique for use in large-scale deep learning pipelines. Additionally, we discussed the potential ethical considerations of using spectrogram flipping and provided guidelines for mitigating the risk of generating biased data.

Overall, our findings demonstrate that spectrogram flipping is a promising technique for improving the performance and robustness of deep learning models in various audio classification tasks. We hope that our research will inspire further exploration and experimentation with this technique in the field of audio signal processing and deep learning.

REFERENCES

[1]. Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. In Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH) (pp. 2613-2617).

[2]. Salamon, J., Jacoby, C., & Bello, J. P. (2017). A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM International Conference on Multimedia (pp. 1041-1049).

[3]. Mauch, M., Dixon, S., & Sandler, M. (2016). Exploring data augmentation for improved singing voice detection with neural networks. In Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR) (pp. 291-297).

[4]. Mesaros, A., Heittola, T., Benetos, E., Foster, P., Lagrange, M., Virtanen, T., & Plumbley, M. D. (2018). On the use of data augmentation for acoustic event detection using deep learning. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (pp. 1-5).

[5]. Kumar, A., Kumar, A., Aggarwal, A., & Garg, K. (2019). Speech data augmentation for deep learning. In Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5).

[6]. Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

[7]. Choi, K., Lee, S., & Lee, S. (2020). Mixup-based audio data augmentation for speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7059-7063).

[8]. Abunasser, B. S. Daud,S. M., Zaqout, I., Abu-Naser S. S. Abunaser - A Novel Data Augmentation Algorithm For Datasets With Numerical Features. Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 11.

[9]. Alrakhawi, H. A., Jamiat, N., Umar, I. N., Abu-Naser, S. S. Improvement of Students Achievement by Using Intelligent Tutoring Systems - A Bibliometric Analysis and Reviews. Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 11.

[10]. Abunasser, B.S., Al-Hiealy, M.R.J., Zaqout, I.S., Abu-Naser, S.S. Convolution Neural Network for Breast Cancer Detection and Classification Using Deep Learning. Asian Pacific journal of cancer prevention: APJCP, 2023, 24(2), pp. 531–544

[11]. Alrakhawi, H. A., Jamiat, N., Abu-Naser, S. S. Intelligent Tutoring Systems in Education: A Systematic Review of Usage, Tools, Effects and Evaluation. Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 4, pp. 1205-1226.

[12]. Zarandah, Q. M. M., Daud, S. M., Abu-Naser, S. S. A Systematic Literature Review Of Machine and Deep Learning-Based Detection And Classification Methods for Diseases Related To the Respiratory System, Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 4, pp. 1273-1296.

[13]. Alkayyali, Z. K. D., Idris, S. A. B, Abu-Naser, S. S. A Systematic Literature Review of Deep and Machine Learning Algorithms in Cardiovascular Diseases Diagnosis, Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 4, pp. 1353-1365.

[14]. Abunasser, B. S. Daud,S. M., Zaqout,I., Abu-Naser S. S. Convolution Neural Network For Breast Cancer Detection And Classification - Final Results. Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 1, pp. 315-329.

[15]. Taha, A. M. H., Ariffin, D. S. B. B., Abu-Naser, S. S. A Systematic Literature Review of Deep and Machine Learning Algorithms in Brain Tumor and Meta-Analysis, Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 1, pp. 21-36.

[16]. Abu Ghosh, M.M., Atallah, R.R., Abu Naser, S.S. Secure mobile cloud computing for sensitive data: Teacher services for palestinian higher education institutions. International Journal of Grid and Distributed Computing, 2016, vol. 9, no. 2, pp. 17–22

[17]. Abu Naser, S.S. Evaluating the effectiveness of the CPP-Tutor, an intelligent tutoring system for students learning to program in C++. Journal of Applied Sciences Research, vol. 5, no. 1, pp. 109-114, 2009.

[18]. Abu-Naser, S.S., El-Hissi H., Abu-Rass, M., & El-khozondar, N. An expert system for endocrine diagnosis and treatments using JESS. Journal of Artificial Intelligence; vol. 3, no. 4, pp. 239-251, 2010.

[19]. Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S. and Abu-Naser, S. S. "Breast Cancer Detection and Classification using Deep Learning Xception Algorithm" International Journal of Advanced Computer Science and Applications(IJACSA), 13(7),223-228, 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130729

[20]. Abunasser, B.S., AL-Hiealy, M.R. J., Barhoom, A. M. Almasri A. R. and Abu-Naser, S. S. "Prediction of Instructor Performance using Machine and Deep Learning Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 13(7), 78-83, 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130711

[21]. Alayoubi, M.M., Arekat, Z.M., Al Shobaki, M.J., Abu-Naser, S.S. The Impact of Work Stress on Job Performance Among Nursing Staff in Al-Awda Hospital. Foundations of Management, 2022, 14(1), pp. 87–108

[22]. Albatish, I.M., Abu-Naser, S.S. Modeling and controlling smart traffic light system using a rule based system. Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019, 2019, pp. 55–60, 8925318

[23]. Almasri, A., Obaid, T., Abumandil, M.S.S., ...Mahmoud, A.Y., Abu-Naser, S.S. Mining Educational Data to Improve Teachers' Performance. Lecture Notes in Networks and Systems, 2023, 550 LNNS, pp. 243–255

[24]. Almasri, A.R., Yahaya, N.A., Abu-Naser, S.S. Instructor Performance Modeling For Predicting Student Satisfaction Using Machine Learning - Preliminary Results. Journal of Theoretical and Applied Information Technology, 2022, 100(19), pp. 5481–5496

[25]. Alsharif, F. Safi S., AbouFoul T., Abu Nasr, M., Abu Nasser S. Mechanical Reconfigurable Microstrip Antenna. International Journal of Microwave and Optical Technology, vol. 11, no. 3, pp.153-160, 2016.

[26]. Arqawi, S., Atieh, K.A.F.T., Shobaki, M.J.A.L., Abu-Naser, S.S., Abu Abdulla, A.A.M. Integration of the dimensions of computerized health information systems and their role in improving administrative performance in Al-Shifa medical complex, Journal of Theoretical and Applied Information Technologythis link is disabled, 2020, vol. 98, no. 6, pp. 1087–1119

[27]. Arqawi, S.M., Abu Rumman, M.A., Zitawi, E.A., ...Abunasser, B.S., Abu-Naser, S.S. Predicting Employee Attrition And Performance Using Deep Learning. Journal of Theoretical and Applied Information Technology, 2022, 100(21), pp. 6526–6536

[28]. Arqawi, S.M., Zitawi, E.A., Rabaya, A.H., Abunasser, B.S., Abu-Naser, S.S., "Predicting University Student Retention using Artificial Intelligence", International Journal of Advanced Computer Science and Applications , 2022, vol. 13, no. 9, pp. 315–324

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[29]. Barhoom, A.M.A., Al-Hiealy, M.R.J., Abu-Naser, S.S. Bone Abnormalities Detection and Classification Using Deep Learning-VGG16 Algorithm. Journal of Theoretical and Applied Information Technology, 2022, 100(20), pp. 6173–6184

[30]. Barhoom, A.M.A., Al-Hiealy, M.R.J., Abu-Naser, S.S. Deep Learning-Xception Algorithm for Upper Bone Abnormalities Classification. Journal of Theoretical and Applied Information Technology, 2022, 100(23), pp. 6986–6997

[31]. Buhisi, N. I., & Abu-Naser, S. S. Dynamic programming as a tool of decision supporting. Journal of Applied Sciences Research. Vo. 5, no. 6, pp. 671-676, 2009.

[32]. El-Habil, B.Y., Abu-Naser, S.S. Global Climate Prediction Using Deep Learning. Journal of Theoretical and Applied Information Technology, 2022, 100(24), pp. 4824–4838

[33]. Elzamly, A., Hussin, B., Naser, S.A., ...Selamat, A., Rashed, A. A new conceptual framework modelling for cloud computing risk management in banking organizations. International Journal of Grid and Distributed Computing, 2016, vol. 9, no. 9, pp. 137–154

[34]. Elzamly, A., Messabia, N., Doheir, M., ...Al-Aqqad, M., Alazzam, M. Assessment risks for managing software planning processes in information technology systems. International Journal of Advanced Science and Technology, 2019, vol. 28, no. 1, pp. 327–338

[35]. Eneizan, B., Obaid, T., Abumandil, M.S.S., ...Arif, K., Abulehia, A.F.S. Acceptance of Mobile Banking in the Era of COVID-19. Lecture Notes in Networks and Systems, 2023, 550 LNNS, pp. 29–42

[36]. Alzamily, J. Y. I., Ariffin, S. B., Abu-Naser, S. S. Classification of Encrypted Images Using Deep Learning –Resnet50. Journal of Theoretical and Applied Information Technology, 2022, 100(21), pp. 6610–6620

[37]. Mady, S.A., Arqawi, S.M., Al Shobaki, M.J., Abu-Naser, S.S. Lean manufacturing dimensions and its relationship in promoting the improvement of production processes in industrial companies. International Journal on Emerging Technologies, 2020, vol. 11, no. 3, pp. 881–896

[38]. Naser, S. S. A. Developing an intelligent tutoring system for students learning to program in C++. Information Technology Journal, vol. 7, no. 7, pp. 1051-1060, 2008.

[39]. Naser, S. S. A. Developing visualization tool for teaching AI searching algorithms. Information Technology Journal, vol. 7, no. 2, pp. 350-355, 2008.

[40]. Naser, S. S. A. Intelligent tutoring system for teaching database to sophomore students in Gaza and its effect on their performance. Information Technology Journal, vol. 5, no. 5, pp. 916-922, 2006.

[41]. Naser, S. S. A. JEE-Tutor: An intelligent tutoring system for java expressions evaluation. Information Technology Journal, vol. 7, no. 3, pp. 528-532, 2008.

[42]. Obaid, T., Eneizan, B., Naser, S.S.A., ...Abualrejal, H.M.E., Gazem, N.A. Factors Contributing to an Effective E- Government Adoption in Palestine. Lecture Notes on Data Engineering and Communications Technologies, 2022, 127, pp. 663–676

[43]. Obaid, T., Eneizan, B., Abumandil, M.S.S., ...Abu-Naser, S.S., Ali, A.A.A. Factors Affecting Students' Adoption of E-Learning Systems During COVID-19 Pandemic: A Structural Equation Modeling Approach. Lecture Notes in Networks and Systems, 2023, 550 LNNS, pp. 227–242

[44]. Saleh, A., Sukaik, R., Abu-Naser, S.S. Brain tumor classification using deep learning. Proceedings - 2020 International Conference on Assistive and Rehabilitation Technologies, iCareTech 2020, 2020, pp. 131–136, 9328072