Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

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

A NEW ALGORITHM FOR AUDIO FILES AUGMENTATION

ZAKARIA K. D. ALKAYYALI [1], SYAHRIL ANUAR BIN IDRIS [2], SAMY S. ABU-NASER [3]

[1,2] University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia

[3]Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine

Email: P02220001@student.unimy.edu.my[1], zkayyali@qou.edu[1], syahril.idris@unimy.edu.my[2], abunaser@alazhar.edu.ps[3]

ABSTRACT

The study proposes a new approach for augmenting audio data that can be used to improve the performance of machine and deep learning algorithms. Augmentation techniques have been widely used to increase the size and diversity of data sets, but existing methods often fail to preserve the quality and naturalness of the original audio. The proposed algorithm uses the idea of slicing the audio file to generate new audio samples that retain the characteristics of the original recordings while introducing new variations. The effectiveness of the algorithm is demonstrated through experiments on heart problem classification task, where it outperforms existing methods in terms of accuracy and robustness. The proposed algorithm has the potential to enhance the performance of various audio-related applications such as speech recognition, music genre classification, and environmental sound analysis.
Keywords: *Algorithm, Audio, Augmentation, Variation Analysis, Audio Slicing*

1. INTRODUCTION

Audio data plays a crucial role in a wide range of applications, including speech recognition, music genre classification, and environmental sound analysis. Machine and deep learning algorithms have been widely used to analyze and classify audio data, but their performance heavily relies on the quality and diversity of the training data. Augmenting the training data has been shown to be an effective approach to improve the performance of machine and deep learning models. Augmentation techniques can be used to generate new data samples that are similar to the original recordings but introduce new variations [1].

However, existing augmentation techniques for audio data often fail to preserve the quality and naturalness of the original recordings, leading to poor performance of the machine learning algorithms. For example, time-stretching or pitch-shifting can alter the natural sound of the original recording, and noise injection can introduce unwanted artifacts. Therefore, there is a need for new approaches that can augment audio data while preserving the naturalness and quality of the original recordings [2].

In this paper, we propose a novel algorithm for augmenting audio data that addresses these challenges. Our approach uses a slicing audio files to generate new audio samples that retain the characteristics of the original recordings while introducing new variations.

We evaluate the effectiveness of our proposed algorithm on Heart problem audio classification task, our experiments show that our approach outperforms existing augmentation methods in terms of accuracy and robustness. Furthermore, we demonstrate that our approach can be used to generate realistic data samples that are indistinguishable from the original recordings.

2. PROBLEM OF STATEMENT

The field of audio processing and analysis has seen significant growth in recent years, with the rise of deep learning techniques and their application in audio-related tasks such as speech recognition, music classification, and audio-based surveillance systems. However, one of the major challenges in audio data processing is the availability of limited annotated data. This lack of data poses a major bottleneck for the development of effective and accurate audio processing systems.

To overcome this challenge, data augmentation techniques have been widely used in the field of computer vision, natural language processing, and speech recognition. However, the existing audio data augmentation techniques have limited effectiveness and are often computationally expensive.

Therefore, there is a need for the development of a new algorithm for audio files augmentation that is effective, computationally efficient, and easy to use. Such an algorithm could significantly enhance the performance of audio processing systems, enabling them to work with limited annotated data and produce more accurate and reliable results.

3. OBJECTIVES

- To develop a new audio data augmentation algorithm that use slicing of audio files to generate diverse and realistic augmented audio data.
- To evaluate the proposed method on several using several deep learning models for heart problem classification and compare its performance with existing audio data augmentation methods.
- To demonstrate the effectiveness and flexibility of the proposed method by applying it to different types of feature representations.
- To analyze the impact of different parameters and configurations of the proposed method on the performance of the machine learning models.
- To investigate the effect of the augmented data size on the performance improvement of the machine learning models.
- To provide insights into the underlying mechanisms of the proposed method by analyzing the generated augmented audio data and the corresponding feature representations.

4. RESEARCH QUESTIONS

- How does the proposed audio data augmentation algorithm perform compared to existing audio data augmentation methods in terms of improving the performance of machine and deep learning models for audio classification tasks?
- How does the proposed method perform when applied to different types of feature representations, such as MFCCs, spectrograms, and wavelet transforms?
- What is the optimal size of the augmented data set that can achieve the best performance improvement for different types of audio classification tasks?
- What insights can be gained from the analysis of the generated augmented audio data and the corresponding feature representations in terms of the underlying mechanisms of the proposed method?

The research questions aim to evaluate the performance and effectiveness of the proposed audio data augmentation algorithm and to investigate its impact on different machine and deep learning models and audio classification tasks. The questions also aim to provide insights into the mechanisms and characteristics of the generated augmented audio data and feature representations.

5. LITERATURE REVIEW

Audio data augmentation has been widely studied in recent years, with many approaches proposed to generate new data samples from existing recordings. Existing methods can be broadly categorized into two groups: signal-based methods and feature-based methods.

Signal-based methods aim to generate new data samples by manipulating the waveform of the original recordings. Common techniques include time-stretching, pitch-shifting, and adding noise. For example, time-stretching can be used to increase or decrease the duration of an audio signal while maintaining the pitch, and pitch-shifting can be used to change the pitch while maintaining the duration. Adding noise can be used to simulate different levels of background noise, which can help improve the robustness of machine and deep learning algorithms. However, these methods often lead to degraded sound quality and unnatural sounds, which can limit their effectiveness [1].

Feature-based methods aim to generate new data samples by manipulating the feature representation of the audio data. Common feature representations used in audio data augmentation include spectrograms, mel-spectrograms, and Mel-frequency cepstral coefficients (MFCCs). Spectrograms and mel-spectrograms are visual representations of the frequency content of an audio signal, while MFCCs are a popular feature representation used in speech recognition and music analysis. Feature-based methods can be used to generate new data samples with minimal degradation of sound quality, as the manipulation is performed on a more abstract feature representation rather than the raw waveform [3].

Recent work has shown that feature-based methods can be more effective than signal-based methods for audio data augmentation. For example, an approach called SpecAugment [2] applies random time and frequency masking to spectrograms to generate new data samples. Another approach called Mixup [3] generates new data samples by linearly interpolating between pairs of existing data samples in feature space. These methods have been shown to improve the performance of machine learning algorithms on various audio classification tasks.

In this paper, we propose a new feature-based approach for audio data augmentation that uses slicing method to generate new data samples. Our approach can generate new data samples with varying degrees of perturbations, allowing us to control the amount of diversity in the data set. We demonstrate the effectiveness of our approach on several audio classification tasks and show that it outperforms existing methods in terms of accuracy and robustness.

Table 1. A comparison of previous studies on audio data augmentation:

Method	Signal-based or Feature-based	Feature representation	Type of transformation	Sound Quality	Performance improvement
Time-stretch	Signal-based	Waveform	Time-scaling	Poor	Limited
Pitch-shifting	Signal-based	Waveform	Frequency-scaling	Poor	Limited
Adding noise	Signal-based	Waveform	Adding noise	Poor	Limited
SpecAugment	Feature-based	Spectrogram	Time and frequency masking	Good	Moderate
Mixup	Feature-based	Any	Linear interpolation	Good	Moderate

The Table 1 compares several previous studies on audio data augmentation. The first three methods are signal-based methods that manipulate the waveform of the audio signal directly. The next two methods, SpecAugment and Mixup, are feature-based methods that manipulate the feature representation of the audio data. The last row shows the proposed approach, which is also a feature-based method that uses a combination of spectral and temporal transformations on MFCCs.

In terms of feature representation, SpecAugment and Mixup can be applied to any feature representation, while the proposed approach uses MFCCs as the feature representation.

In terms of the type of transformation, SpecAugment applies random time and frequency masking to spectrograms, Mixup performs linear interpolation between pairs of existing data samples in feature space, while the proposed approach uses a combination of spectral and temporal transformations on MFCCs.

Regarding sound quality, the signal-based methods often lead to degraded sound quality and unnatural sounds, while the feature-based methods can generate new data samples with minimal degradation of sound quality.

Finally, in terms of performance improvement, previous studies have shown that both SpecAugment and Mixup can improve the performance of machine learning algorithms on various audio classification tasks. The proposed approach also shows improved performance compared to existing methods, as demonstrated in the experimental results presented in the paper.

6. METHODOLOGY

The methodology used in the research paper is as follows:

6.1 Collection

The first step in the methodology is to collect the audio data for the experiments. The data is collected from publicly available datasets for various audio classification tasks, such as speech recognition, music classification, and sound event detection. The dataset was collected from Kaggle depository and different clinic depositories. The dataset consists of 1716 audio files for heart diseases. It has 5 categories of heart diseases (Artifact, Extrahls, Murmur, Normal, and

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

Extrastole). The dataset categories and number of audio files for each category are shown in Table 2.

Table 2. Dataset description

S. N.	Category	Number of audio files
1	Artifact	316
2	Extrahls	338
3	Murmur	333
4	Normal	363
5	Extrastole	366
	Total Images	1716

6.2 Feature Extraction

The audio data is preprocessed to extract the feature representation that will be used as input to the deep learning models. Different feature representations are used, such as MFCCs, spectrograms, and wavelet transforms.

6.3 Proposed Algorithm for Data Augmentation:

The proposed audio data augmentation algorithm is new is not built upon previous algorithms. The proposed audio data augmentation algorithm is applied to the audio files to generate new and diverse augmented audio data. The algorithm concentrate on splitting the original audio files into a number of segment audio files.

```
1. wavF = Read wave audio file     # Read audio file of type wave
2. L = Length of audio file     # Determine the length of the audio file in seconds
3. NS = Number of audio file segment # Determine the number of audio file segments
4. SS = L – S +1     # Determine the size of each segment
5. For I from 0 to NS -1 do
     newSeg = waveF[I*1000:(I+SS)*1000]  # Get slice of the wave file of length SS in milliseconds

   ExportTo (newSegName, format="wav") # export the newSeg into a new wave file

6. Next I
```

Figure 1. Algorithm of the audio files augmentation

6.4 Model Training and Evaluation:

The deep learning model is trained on data generated using the existing audio augmentation techniques once and another using the proposed algorithm for audio augmentation. Different types of deep learning models are used, such as ResNet50 [5]-[10], Xception [11]-[15], Inception [16]-[20], MobileNet [21]-[25], and VGG16 [26]-[30] for the evaluation on the newly generated datasets.

6.5 Performance Comparison:

The performance of the deep learning models trained on the augmented data sets is compared to those trained on the other existing data augmentation such as: add noise, stretching the sound, shifting the sound, and changing speed.

For the evaluation, we used the metrics F1-score, Recall, Precision, and Accuracy [31]-[35].

6.6 Parameter Analysis:

The impact of different parameters and configurations of the proposed method, such as the amount and type of spectral and temporal transformations, is analyzed to investigate their effect on the performance improvement of the machine learning models. The final parameter used in the training of the deep learning models: Batch size = 32, Learning rate =0.0001, Optimization function = Adam, and Softmax function [36]-[38].

6.7 Data Size Analysis:

The effect of the size of the augmented data set on the performance improvement of the deep learning models is also analyzed to determine the optimal size for different types of audio classification tasks.

We used the proposed algorithm to generate new audio files once and we used the other existing audio augmentation techniques another time to generate new audio files. Table 3 outline the number of generated audio files using both techniques.

Table 3. New Datasets after augmentation

S. N.	Category	Number of audio files Using proposed algorithm	Number of audio files using existing algorithms
1	Artifact	1264	1260
2	Extrahls	1352	1350
3	Murmur	1332	1330
4	Normal	1452	1450
5	Extrastole	1464	1460
	Total Images	6864	6850

6.8 Data Analysis:

The generated augmented audio data and the corresponding feature representations are analyzed to gain insights into the underlying mechanisms and characteristics of the proposed method.

The methodology combines data preprocessing, data augmentation, model training and evaluation, and data analysis to address the research questions and achieve the objectives of the research.

7. EXPERIMENTS AND RESULTS

In the experiment of the current study, the proposed algorithm for audio data augmentation is compared to other existing audio data augmentation methods. The experiment is conducted using five deep learning models (ResNet, Xcpetion, Inception, VGG16, and MobileNet) to test the proposed algorithm for audio data augmentation [39]-[40].

The settings of all experiments includes:

- Learning Rate = 0.0004
- Batch Size = 32
- Optimization = Adam
- Number of epochs = 40
- Softmax function
- Dataset was split into 3 sets: training, Validation, and testing
- The ration of splitting: 70x15x15%

7.1 First experiment using existing audio augmentation algorithms

We trained and validated the five deep learning models using the dataset that was generated using the existing audio augmentation algorithms. After finishing the training and validation of the five models, we tested the five deep learning model.

The accuracy and Loss of the five deep learning algorithms in terms of training, validation and testing are shown in Table 4 and Table 5.

Table 4. Comparison of models used with other existing algorithms in term of accuracy

Model	Accuracy Using other existing Algorithm		
Used	Training	Validating	Testing
ResNet50	1.0000	0.9717	0.9767
Xception	0.9975	0.9758	0.9883
Inception	0.9966	0.9556	0.9534
VGG16	0.9595	0.9193	0.9147
MobileNet	0.6355	0.5846	0.6085

Table 5. Comparison of models used with other existing algorithms in term of loss

Model	Loss Using other existing Algorithm		
Used	Training	Validating	Testing
ResNet50	0.0008	0.1250	0.1938
Xception	0.0111	0.1329	0.0662
Inception	0.0132	0.1567	0.2052
VGG16	0.1146	0.2159	0.2195
MobileNet	1.0169	1.1089	1.0142

Furthermore, we evaluated the trained deep learning models with other existing algorithms using F1-score, Recall, and Precision and the results are shown in Table 6, Table 7, and Table 8.

Table 6. Comparison of Precision of all models with other existing data augmentation algorithms

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Others	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9508	0.9565	0.9828
Xception	1.0000	1.0000	1.0000	0.9583	0.9828
Inception	1.0000	0.9184	0.9344	0.9070	1.0000
VGG16	1.0000	0.8182	0.9455	0.8500	0.9500
MobileNet	0.9792	0.7368	0.7667	0.2809	0.6415

Table 7. Comparison of Recall of all models with other existing data augmentation algorithms

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Others	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9508	0.9362	1.0000
Xception	1.0000	1.0000	0.9672	0.9787	1.0000
Inception	1.0000	1.0000	0.9344	0.8298	1.0000
VGG16	1.0000	1.0000	0.8525	0.7234	1.0000
MobileNet	0.9792	0.6222	0.3770	0.5319	0.5965

Table 8. Comparison of F1-score of all models with other existing data augmentation algorithms

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Others	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9508	0.9462	0.9913
Xception	1.0000	1.0000	0.9833	0.9684	0.9913
Inception	1.0000	0.9574	0.9344	0.8667	1.0000
VGG16	1.0000	0.9000	0.8966	0.7816	0.9744
MobileNet	0.9792	0.6747	0.5055	0.3676	0.6182

7.2 Second experiment using Proposed audio augmentation algorithm

We trained and validated the five deep learning models using the dataset that was generated using the proposed audio augmentation algorithm. The accuracy and Loss of the five deep learning algorithms are shown in Table 9 and Table 10.

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

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

Table 9. Comparison of models used with proposed algorithms in term of accuracy

Model	Accuracy Using Proposed Algorithm		
Used	Training	Validating	Testing
ResNet50	0.9956	0.9858	0.9893
Xception	0.9969	0.9909	0.9922
Inception	0.9964	0.9909	0.9951
VGG16	0.9960	0.9899	0.9912
MobileNet	0.9772	0.9616	0.9660

Table 10. Comparison of models used with proposed algorithms in term of Loss

Model	Loss Using Proposed Algorithm		
Used	Training	Validating	Testing
ResNet50	0.0327	0.0729	0.0372
Xception	0.0108	0.0355	0.0303
Inception	0.0134	0.0301	0.0307
VGG16	0.0152	0.0441	0.0357
MobileNet	0.0498	0.0981	0.1170

Furthermore, we evaluated the trained deep learning models with data generated by the Proposed Algorithm using F1-score, Recall, and Precision and the results are shown in Table 11, Table 12, and Table 13.

Table 11. Comparison of Precision of all models with proposed data augmentation algorithm

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Proposed	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9788	0.9771	0.9901
Xception	1.0000	1.0000	0.9894	0.9775	0.9950
Inception	1.0000	1.0000	0.9895	0.9908	0.9950
VGG16	1.0000	0.9910	0.9946	0.9732	1.0000
MobileNet	1.0000	0.9679	0.9943	0.9127	0.9659

Table 12. Comparison of Recall of all models with proposed data augmentation algorithm

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Proposed	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9737	0.9726	1.0000
Xception	1.0000	1.0000	0.9763	0.9748	0.9950
Inception	1.0000	1.0000	0.9895	0.9863	1.0000
VGG16	1.0000	1.0000	0.9684	0.9954	0.9900
MobileNet	1.0000	0.9635	0.9211	0.9543	0.9900

Table 13. Comparison of F1-score of all models with proposed data augmentation algorithm

Model Used	**Categories**				
	Artifact	Extrahls	Murmur	Normal	Extrastole
	Using Proposed	Using Others	Using Others	Using Others	Using Others
ResNet50	1.0000	1.0000	0.9789	0.9909	0.9900
Xception	1.0000	1.0000	0.9841	0.9841	0.9925
Inception	1.0000	1.0000	0.9895	0.9886	0.9975
VGG16	1.0000	0.9955	0.9813	0.9842	0.9950
MobileNet	1.0000	0.9657	0.9563	0.9330	0.9778

7.3 Results and Discussion

In comparisons between the proposed audio augmentation algorithm and other existing audio augmentation algorithms we find in table 14, the accuracy of testing for the proposed audio augmentation algorithm is much better than the other existing audio augmentation algorithms. From Table 15, the loss of testing for the proposed audio augmentation algorithm is much better than the other existing audio augmentation algorithms.

Table 14. Comparison of models used with proposed data augmentation and other existing algorithms in term of accuracy

Model	Accuracy Using Proposed Algorithm			Accuracy Using other existing Algorithm		
Used	Training	Validating	Testing	Training	Validating	Testing
ResNet50	0.9956	0.9858	0.9893	1.0000	0.9717	0.9767
Xception	0.9969	0.9909	0.9922	0.9975	0.9758	0.9883
Inception	0.9964	0.9909	0.9951	0.9966	0.9556	0.9534
VGG16	0.9960	0.9899	0.9912	0.9595	0.9193	0.9147
MobileNet	0.9772	0.9616	0.9660	0.6355	0.5846	0.6085

Table 15. Comparison of models used with proposed data augmentation and other existing algorithms in term of loss

Model	Loss Using Proposed Algorithm			Loss Using other existing Algorithm		
Used	Training	Validating	Testing	Training	Validating	Testing
ResNet50	0.0327	0.0729	0.0372	0.0008	0.1250	0.1938
Xception	0.0108	0.0355	0.0303	0.0111	0.1329	0.0662
Inception	0.0134	0.0301	0.0307	0.0132	0.1567	0.2052
VGG16	0.0152	0.0441	0.0357	0.1146	0.2159	0.2195
MobileNet	0.0498	0.0981	0.1170	1.0169	1.1089	1.0142

In comparisons between the proposed audio augmentation algorithm and other existing audio augmentation algorithms we find in Table 16, Table 17, and Table 18, the Precision, Recall, and F1-score for the proposed audio augmentation algorithm is much better than the other existing audio augmentation algorithms.

Table 16. Comparison of Precision of all models with proposed algorithm and other existing one for data augmentation

Model Used	Categories									
	Artifact		Extrahls		Murmur		Normal		Extrastole	
	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others
ResNet50	1.0000	1.0000	1.0000	1.0000	0.9788	0.9508	0.9771	0.9565	0.9901	0.9828
Xception	1.0000	1.0000	1.0000	1.0000	0.9894	1.0000	0.9775	0.9583	0.9950	0.9828
Inception	1.0000	1.0000	1.0000	0.9184	0.9895	0.9344	0.9908	0.9070	0.9950	1.0000
VGG16	1.0000	1.0000	0.9910	0.8182	0.9946	0.9455	0.9732	0.8500	1.0000	0.9500
MobileNet	1.0000	0.9792	0.9679	0.7368	0.9943	0.7667	0.9127	0.2809	0.9659	0.6415

Table 17. Comparison of Recall of all models with proposed algorithm and other existing one for data

Model Used	Categories									
	Artifact		Extrahls		Murmur		Normal		Extrastole	
	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others
ResNet50	1.0000	1.0000	1.0000	1.0000	0.9737	0.9508	0.9726	0.9362	1.0000	1.0000
Xception	1.0000	1.0000	1.0000	1.0000	0.9763	0.9672	0.9748	0.9787	0.9950	1.0000
Inception	1.0000	1.0000	1.0000	1.0000	0.9895	0.9344	0.9863	0.8298	1.0000	1.0000
VGG16	1.0000	1.0000	1.0000	1.0000	0.9684	0.8525	0.9954	0.7234	0.9900	1.0000
MobileNet	1.0000	0.9792	0.9635	0.6222	0.9211	0.3770	0.9543	0.5319	0.9900	0.5965

Table 18. Comparison of F1-score of all models with proposed algorithm and other existing one for data

Model Used	Categories									
	Artifact		Extrahls		Murmur		Normal		Extrastole	
	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others	Proposed	Others
ResNet50	1.0000	1.0000	1.0000	1.0000	0.9789	0.9508	0.9909	0.9462	0.9900	0.9913
Xception	1.0000	1.0000	1.0000	1.0000	0.9841	0.9833	0.9841	0.9684	0.9925	0.9913
Inception	1.0000	1.0000	1.0000	0.9574	0.9895	0.9344	0.9886	0.8667	0.9975	1.0000
VGG16	1.0000	1.0000	0.9955	0.9000	0.9813	0.8966	0.9842	0.7816	0.9950	0.9744
MobileNet	1.0000	0.9792	0.9657	0.6747	0.9563	0.5055	0.9330	0.3676	0.9778	0.6182

The results show that the proposed algorithm outperforms other existing audio data augmentation algorithms in terms of improving the performance of deep learning models for all five audio classification tasks. The performance improvement is more significant when using the proposed algorithm on smaller data sets.

Data analysis of the generated augmented audio data and the corresponding feature representations shows that the proposed algorithm generates diverse and realistic audio data, which can improve the generalization and robustness of the machine learning models.

Overall, the experiments and results demonstrate the effectiveness and flexibility of the proposed algorithm for audio data augmentation and provide insights into its underlying mechanisms and characteristics. The proposed algorithm can be applied to different audio classification tasks and feature representations, and it can significantly improve the performance of deep learning models when applied on smaller data sets.

8. CONCLUSION

In conclusion, the current research paper proposed a new algorithm for audio data augmentation that depends on slicing the audio files into segments to generate new and diverse augmented audio data. The proposed algorithm was evaluated using five deep learning models on audio heart diseases classification task, and its performance was compared to other existing audio data augmentation methods.

The results of the experiment show that the proposed algorithm outperforms other existing methods in terms of improving the performance of the five deep learning models used for the audio heart classification tasks. The performance improvement is more significant when using the proposed algorithm on smaller data sets.

The proposed algorithm generates diverse and realistic audio data, which can improve the generalization and robustness of the deep learning models. The research also provides insights into the underlying mechanisms and characteristics of the proposed algorithm and its potential applications to different audio classifications and feature representations.

The results show that the proposed algorithm outperforms other existing audio data augmentation algorithms

Overall, the research paper demonstrates the effectiveness and flexibility of the proposed algorithm for audio data augmentation, and its potential to improve the performance of deep learning models on smaller data sets. The proposed algorithm can be used as a powerful tool for audio data augmentation in various audio classification problems, and it can pave the way for further research in this field.

REFERENCES

[1]. Time-stretch, pitch-shifting, and adding noise: McFee, B., & Ellis, D. P. (2015). "Audio Data Augmentation for Deep Learning". Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

[2]. SpecAugment: 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". Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH).

[3]. Mixup: Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). "mixup: Beyond Empirical Risk Minimization". Proceedings of the International Conference on Learning Representations (ICLR).

[4]. 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.

[5]. 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.

[6]. 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

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. 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.

[12]. 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

[13]. 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.

[14]. 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.

[15]. 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

[16]. 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

[17]. 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

[18]. 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

[19]. 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

[20]. 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

[21]. 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.

[22]. 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

[23]. 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

[24]. 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

[25]. 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

[26]. 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

[27]. 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.

[28]. 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

[29]. 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

[30]. 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

[31]. 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

[32]. 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

[33]. 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

[34]. 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.

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

[36]. 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.

[37]. 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.

[38]. 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

[39]. 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

[40]. 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