ENHANCING MACHINE LEARNING MODELS WITH AUDIO AUGMENTATION TECHNIQUES

Andrii Pirko[1], Iryna Boretska[2]

[1] postgraduate student, Ukrainian National Forestry University, Lviv, Ukraine. https://orcid.org/0009-0007-9056-0413. pirko.andrii@nltu.lviv.ua.
[2] associate professor of Computer Science Department, Ukrainian National Forestry University, Lviv, Ukraine. https://orcid.org/0000-0002-6767-104X. boretska@nltu.edu.ua.

Abstract – This paper investigates how audio augmentation techniques improve the classification accuracy of guitar chords. By applying methods such as noise addition, speed modification, and time shifting, a CNN trained on an augmented dataset achieved better results compared to non-augmented data. The findings highlight the importance of task-specific augmentation techniques in enhancing audio analysis

Keywords – ***audio augmentation, machine learning, neural networks, audio signal processing*.**

Audio augmentation techniques have gained significant importance in machine learning, particularly in solving tasks related to audio data. These techniques allow for substantial expansion of training datasets and improve the generalization ability of models by adapting them to various acoustic scenarios [1]. The primary aim of this work is to review and analyze existing audio augmentation methods, assess their impact on neural network models, integrate them with modern architectures, and provide recommendations for applying these methods in real-world projects.

Audio augmentation involves modifying original audio signals using various techniques to create new, altered versions of these signals. Key augmentation techniques include:

1. **Noise addition.** One of the most common methods, where white, colored, or other types of noise are added to the original signal. This technique improves

the model's resilience to real-world conditions, where audio signals are often distorted by external noises.

2. **Volume adjustment.** Manipulating the volume level of the audio signal allows the model to work with signals of varying intensity.
3. **Speed alteration.** This technique changes the duration of the audio without affecting its pitch, helping the model handle signals of different durations.
4. **Time shifting.** Shifting the audio signal forward or backward on the timeline, simulating natural variations in recordings.
5. **Frequency and time masking.** A method proposed in SpecAugment [2], where frequency (time) ranges of the spectrogram are masked, increasing the model's robustness to information loss or gaps.
6. **Distortion and reverb.** These techniques add effects that alter the acoustic properties of the signal, simulating real recording conditions.

Augmentation methods can be classified based on the type of manipulation: time-domain, frequency-domain, or intensity-domain manipulations [3]. Each technique can use different sets of parameters, enabling the creation of numerous variations of a single audio signal.

Augmentation has a crucial role in tasks like speech recognition and sound classification. It allows models to better adapt to real-world scenarios where the quality of audio signals can vary significantly [4].

Sound classification, such as recognizing musical instruments or sound events, is an essential task in audio analytics, which can be significantly enhanced by audio augmentation techniques. Audio signal augmentation, which includes speed changes, time shifting, noise addition, and other processing methods, enables models to become more resilient to data variations [5]. This is especially important in real recordings, where audio signals may have different characteristics due to environmental factors, hardware, or even the way sound events are performed.

In the context of guitar chord classification, which can vary greatly depending on the type of instrument, playing style, acoustic conditions, and other factors, applying audio augmentation is critical for improving the reliability and accuracy of

classification models. Using augmentation methods provides better generalization, allowing the model to successfully handle complex and non-standard audio data encountered in real-world situations.

Examples of spectrograms and oscillograms of the C major chord audio signal after applying audio augmentation methods are shown in Figures 1 – 2. These examples illustrate how different augmentation methods alter the original signal and how such changes can affect the classification process.

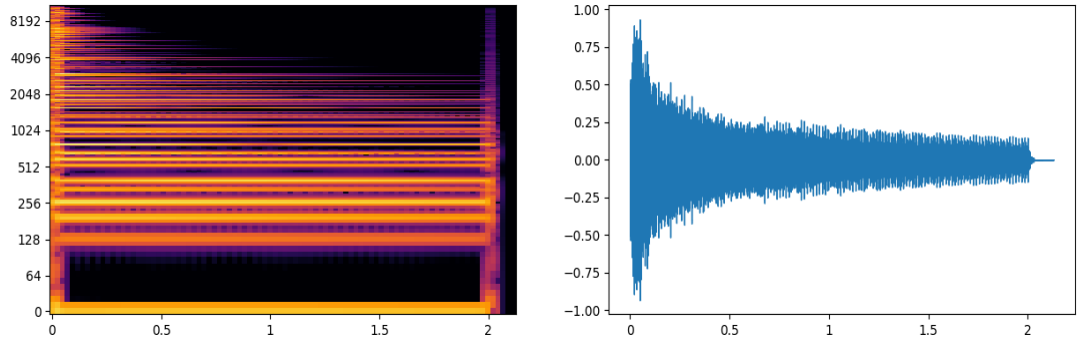

Fig. 1. Spectrogram and oscillogram of the C major chord without audio augmentation

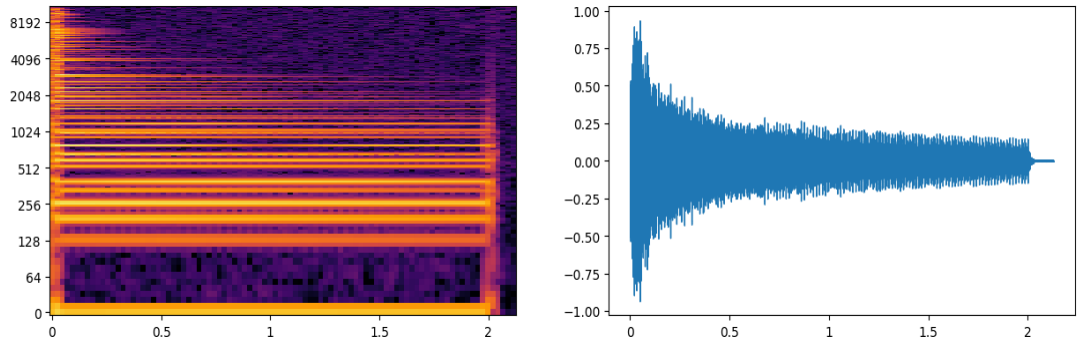

Fig. 2. Spectrogram and oscillogram of the C major chord with added white noise

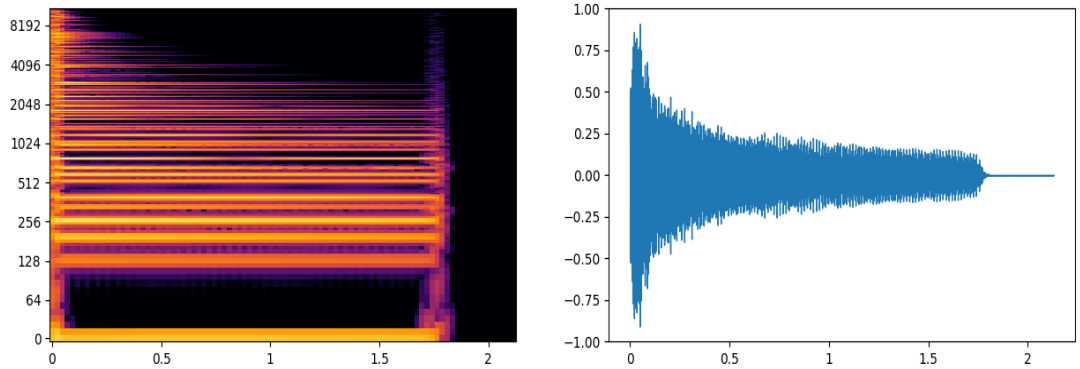

Fig. 3. Spectrogram and oscillogram of the C major chord with speed alteration

Table 1 demonstrates the difference in model training results before and after applying audio augmentation methods.

Table 1

Number of training epochs	Model accuracy on training data without audio augmentation	Model accuracy on test data without audio augmentation	Model accuracy on training data with audio augmentation	Model accuracy on test data with audio augmentation
10	60%	20%	84%	72%
100	87%	32%	96%	81%
1000	92%	27%	97%	88%

Conclusion. Audio augmentation is an essential tool in modern machine learning systems for audio processing. Techniques like noise addition, pitch alteration, and time-frequency masking enhance model robustness to distortions and improve accuracy, especially when original data is limited. The choice of specific augmentation methods depends on the task, with noise addition being effective for speech recognition, and speed and duration alterations beneficial for musical applications. Future advancements in neural network architectures and generative models, such as GANs, will further enhance the capabilities of audio augmentation, creating more adaptive and reliable models.

References

[1] Lucas Ferreira-Paiva, Elizabeth Alfaro-Espinoza, Vinicius M Almeida, Leonardo B Felix, and Rodolpho VA Neves. A survey of data augmentation for audio classification. 2021.

[2] Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, Quoc V. Le (2019). "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition." *Proceedings of the Annual Conference of the International Speech Communication Association*.

[3] Connor Shorten, Taghi M Khoshgoftaar, and Borko Furht. Text data augmentation for deep learning. Journal of big Data, 8:1–34, 2021.

[4] Jinyu Li. Recent advances in end-to-end automatic speech recognition. APSIPA Transactions on Signal and Information Processing, 11(1), 2022.

[5] Dan Oneata and Horia Cucu. Improving multimodal speech recognition by data augmentation and speech representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4579–4588, 2022.