Journal of Engineering

journal homepage: www.jcoeng.edu.iq

Volume 31 Number 4 April 2025

Deep Learning for Lossless Audio Compression

Ali A. Obaid *, **Hasan M. Kadhim**

Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq

ABSTRACT

Audio and speech compression techniques are used to reduce the storage of these data in the required space and the transmission rate of these data in the communication and network systems. In this paper, the researchers exploit neural networks and artificial intelligence to compress audio signals. The researchers investigated compression ratios of 8, 4, 2, and 1 (no compression), and then chose the highest ratio of 8. The compromising choice is based on the best SNR of the recovered audio signal and the required time for implementation. The researchers tested 119 different audio files from the standard BBC audio library. The duration of these files is about 1000 seconds. The average SNR was 26.33 dB, and the mean square error was -52.58 dB. To reduce the running time, the epochs were 30, the hidden layers were 64 to 128, the quantization level was 1, the dimensions were 15 to 20, and each second of the input signal needed 100 seconds to be compressed. The input audio signal files were single-channel mono audio, and the stereo multi-channel audio files were reformatted to mono single-channel. According to the results, the proposal process accomplished good audio compression, while the other parameters were acceptable.

Keywords: Deep learning, Audio compression, Variational autoencoders, Compression ratio.

1. INTRODUCTION

In the digital age, the proliferation of audio content across various platforms has necessitated the development of efficient compression techniques to facilitate storage, transmission, and streaming without compromising the listening experience. Traditional compression methods have long been categorized into two distinct lossy and lossless categories. Lossy compression, exemplified by formats like MP3 and AAC, reduces file size by discarding audio data deemed inaudible or irrelevant to human perception. While this approach has been widely adopted for consumer applications due to its high compression ratios, it inherently results in a loss of audio fidelity, which can be unacceptable for professional audio applications and audiophiles. The technical differences between the two

*Corresponding author

Peer review under the responsibility of University of Baghdad.

https://doi.org/10.31026/j.eng.2025.04.07

Article received: 05/07/2024
Article revised: 27/08/2024
Article accepted: 29/08/2024
Article published: 01/04/2025

types were discussed in detail in the well-known lectures on data, image, and audio lossless and lossy compressions. Lossy compression is favored when the primary concern is file size and the slight loss in quality is acceptable. This is the go-to method for most consumer media applications due to its balance of quality and efficiency **(Jain and Patel, 2009)**. In contrast, lossless compression is the preferred choice when absolute fidelity is non-negotiable. It is used in professional settings where even minor alterations in the data can have significant consequences, such as in music mastering, film post-production, and medical imaging Lossy compression offers the advantage of high compression ratios at the expense of data fidelity, while lossless compression ensures data integrity but with less efficient compression. The decision between the two methods hinges on the balance between quality, storage, and transmission needs **(Jain and Patel, 2009)**.

Conversely, a lossless compression field (such as Apple Lossless Audio Codec (ALAC) and Free Lossless Audio Codec (FLAC)) preserves every bit of the original audio data, ensuring that the decompressed file is identical to the source. However, the trade-off is a more modest reduction in file size, limiting its practicality for applications with stringent storage or bandwidth constraints. The main parameters of lossless audio compression are The Compression Ratio (CR), required time for the compression and decompression process, quality of the output compressed audio, stability, type of compression, threshold, attack, and release. Mainstream audio compression applications and researchers usually compromise among the above parameters **(Crocco et al., 2016; Välimäki and Reiss, 2016)**.

To enhance efficiency and maintain audio quality. **(Shukla et al., 2022)** explore RNNs, CNNs, and GANs for compact audio representations, while **(Dubois et al., 2021)** emphasize cognitive sincerity in compression. **(Barman et al., 2022)** integrate machine learning for adaptive lossless compression, and **(Shukla et al., 2019)** combine DCT and LZW encoding for improved compression rates. **(Hennequin et al., 2017)** propose CNN-based encoding-independent compression techniques. **(Schuller et al., 2002)** develop predictive coding methods to reduce delay and redundancy. **(Ramesh and Wang, 2021)** address real-time audio streaming challenges, and **(Friedland et al., 2020)** analyze perceptual compression's impact on deep learning.

Our research proposes a possible paradigm shift in data compression; utilizing Variational Autoencoders (VAEs) to bridge the gap between lossy and lossless compression. VAEs for audio, language, text, and image processing are efficient deep-learning Digital Signal Processing models that have a high ability to perform different representations for the data of these fields in the future. For instance, by feeding observation input audio signals and data into the VAEs system, a closed and packed latent space can change and configure the observation and machine learning of their expression and presentation. The representations of those latent spaces can ideally extract the principles of data which we request to reformat the high definitions of speech and audio signals and data. Technically, this research is implemented particularly to reduce the size of the lossless compressed audio signal size and rate, i.e. and it can improve the Compression Ratio CR of those observation signals and data of audio and speech. The quality and the fidelity of the processes are acceptable and efficient. The scheme design of accurate VAE performed the above processing when variables of the latent spaces were discretized and then approximated to the best quantization levels. The quality of the lossless output compressed audio was immaculate with a high ratio of compression. That achievement could be evaluated as a remarkable challenge for the compression area. The researchers will describe their technique and how to implement that technique, and then test their results subjectively and objectively of the lossless compression technique based on the VAE algorithms. The researchers will present the details of the analysis for their system about the compression with the results of the ratio, and the quality using different listeners to test the system subjectively. Object tests are used to cross-check these results and to confirm the evaluation of the system. This VAE approach to research can preface the road for other researchers to continue exploiting the efficient ability for deep learning in the audio and speech DSP.

2. VARIATIONAL AUTOENCODERS (VAEs)

Variational AutoEncoders VAEs are a type of rich DSP model in the research of Deep Learning area. More and more researchers are exploiting the efficient performance of Variational Autoencoders due to their abilities for machine learning of the complex statistical distributions of signals and data **(Hemmer et al., 2020; Défossez et al., 2022)**. In speech, audio, text, and image processing, VAEs principally adapt jobs with structured data and multi-dimensional processing. The principles and basics of autoencoders are used in kernels of the VAEs. For the input observation of audio data and signals, their encoding process exploits the Neural Networks (NNs), which have been developed specifically for that. Then, the outputs of the NNs are processed by a latent space formulation. The latent space formulation can inversely decode NN outputs to produce its plaintext original audio and speech data and signals. For generative modeling and/or representation learning, the VAEs are based on statistics, stochastic, and probability analyses. Those analyses gave power to the VAEs. **Fig. 1** describes the scheme of Vibrational AutoEncoders VAEs in sequential stages. The VAE scheme contains an encoder network of the input signal at first and a decoder network of the signal at last. The model maps the input data/ signal to the latent that depends on the statical specification (Probability Distribution Function PDF), which is typically a multivariate Gaussian. This mapping is achieved through sequential Neural Network Layers (NNLs), which process the input mean vector and a variance vector, which together define the parameters of the Gaussian distribution in the latent space. This mapping is also achieved for the output signal/ data. The key here is the introduction of a variational inference framework, which allows the VAE to learn a distribution over latent variables rather than a single deterministic encoding **(Hemmer et al., 2020; Pollastro et al., 2023)**. The decoder, or generative model, takes samples from this learned latent distribution and maps them back into the original data space. The decoder is also a neural network that reconstructs the input data from the latent representation, aiming to minimize the difference between the original and reconstructed data. The training of a VAE involves optimizing a loss function that balances two objectives: the reconstruction loss, which measures the fidelity of the decoded data to the original input, and the divergence of the Kullback-Leibler (KL) algorithm, which makes specific levels (quantifies) for the difference of the chosen prior PDF (usually Normal Gaussian Distribution) and the PDF of the learned latent. This balance ensures that the latent space is both informative about the input data and is regularized to prevent overfitting and to encourage meaningful structure **(San Martin et al., 2019; Kalinin et al., 2021)**.

The use of VAEs extends beyond simple data compression and reconstruction. They have many other uses such as in anomaly detection, where the learned latent representation can detect deviations from the normal data distribution; data generation, where samples from the latent space can be decoded into realistic data instances; and data interpolation, where transitions between different points in the latent space can reveal smooth, interpretable transformations in the data space. The positive merits of VAEs are **(Shang et al., 2021; Dewangan and Maurya, 2021; Cunha et al., 2023)**:

1. VAEs present basic methods for latent representation learning. Learning is available in the generative and the meaningful stages.
2. Due to the above statistical backgrounds, the VAEs provide uncertainty against the robust inference and the approximation of the quantification.
3. With inherently scalable procedures, VAEs can handle complex dataset applications with large-scale ranges.

VAEs can incorporate different designs of other architectures. They have acknowledgments for audio encoding and decoding **(Shang et al., 2021; Dewangan and Maurya, 2021; Cunha et al., 2023)**.

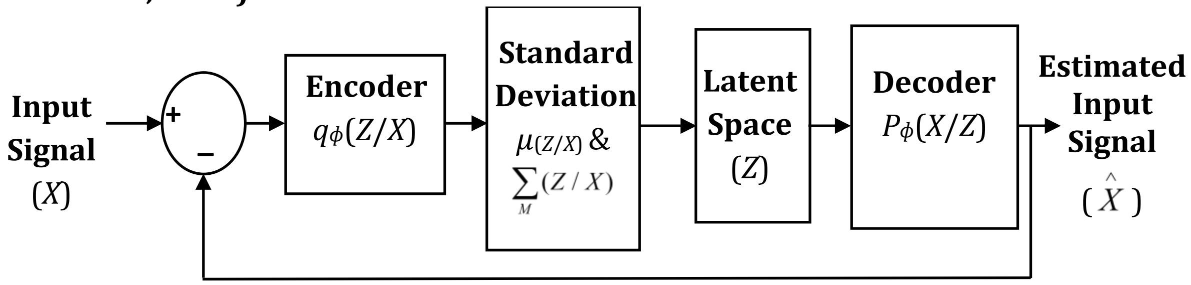
Input Signal (X)
+
−
Encoder $q\phi(Z/X)$
Standard Deviation $\mu_{(Z/X)}$ & $\sum_M (Z/X)$
Latent Space (Z)
Decoder $P\phi(X/Z)$
Estimated Input Signal $(\hat{X})$

Figure 1. Variational Autoencoders Architecture

3. PROPOSED ALGORITHM

The proposed methodology for lossless audio compression using Variational Autoencoders (VAEs) is a systematic approach that combines deep learning techniques with signal processing to achieve high compression ratios without compromising audio fidelity. The core of this methodology lies in the design and training of a VAE tailored to the nuances of audio data, ensuring that the learned latent representation captures the essential features of the audio signal while being amenable to efficient encoding and decoding. The VAE model employed in this work consists of an encoder network and a decoder network, both of which are composed of fully connected layers with rectified linear unit (ReLU) activations. The encoder network takes as input the raw audio waveform, which is typically a one-dimensional time series, and maps it to a lower-dimensional latent space. The encoder outputs two vectors: a mean vector, μ, and a log variance vector, log(σ), which together parameterize a Gaussian distribution from which the latent representation Z is sampled. Mathematically, this process is represented as follows **(Amada et al., 2018; Liu, 2021)**:

$$Z \sim \mathrm{N}\left(\mu, \mathrm{diag}(\sigma^2)\right) \tag{1}$$

Where N denotes the normal distribution, μ is the mean, σ is the variance, and $diag$ indicates a diagonal matrix with the variances along the diagonal. The decoder network, conversely, takes samples from the latent space and reconstructs the original audio waveform. The architecture mirrors that of the encoder, with fully connected layers that progressively increase the dimensionality of the data until it matches the original audio signal. The decoder's output is a reconstruction of the input audio, denoted as x̂, which is compared to the original input (x), to compute the reconstruction loss **(Huang et al., 2019)**. The training of the VAE involves optimizing a loss function that comprises two components: the reconstruction loss (L_{rec}), and the Kullback-Leibler (KL) divergence loss (L_{KL}). These losses are dimensionless (unitless) parameters. The reconstruction loss measures the difference between the original audio and the reconstructed audio, typically using the mean squared error (MSE) for regression tasks **(Yoshimura et al., 2018; Passricha and Aggarwal, 2019)**:

$$L_{rec} = \frac{1}{M} \sum_{i=1}^{i=M} (X_i - \hat{x}_i)^{2} \tag{2}$$

Where M is the number of samples in the batch, and the summation is over all samples.
The KL divergence loss quantifies the dissimilarity between the learned latent distribution and a prior distribution, which is usually a standard normal distribution **(Zeghidour et al., 2021)**:

$$L_{KL} = \frac{1}{2} \sum_{j=1}^{j=M} \left(\mu_j^2 - \sigma_j^2 - \log\left(\sigma_j^2\right) - 1\right) \tag{3}$$

Where the summation is over the dimensions of the latent space.
The total loss (L_{total}) is a weighted sum of the reconstruction loss and the KL divergence loss **(Nagaraj et al., 2020; Nogales et al., 2023)**:

$$L_{total} = L_{rec} + \lambda L_{KL} \tag{4}$$

Where λ is a hyperparameter that controls the trade-off between the two components.
During training, the VAE is optimized using the Adam optimizer, which adjusts the learning rate adaptively for each parameter. The training process involves iteratively updating the weights of the encoder and decoder networks to minimize the total loss, thereby learning a latent space that is both informative and regularized **(Jing et al., 2014; Chen et al., 2021)**.
Once trained, the VAE can be used for compression by encoding the audio into the latent space, which is then quantized to further reduce the data size. The quantized latent representation is losslessly compressed using standard compression algorithms, such as Huffman coding or arithmetic coding. The compressed data can be stored or transmitted, and upon receipt, it is decompressed and decoded using the VAE's decoder network to reconstruct the original audio with no loss in quality **(Shin et al., 2022)**.
The proposed methodology leverages the power of VAEs to learn a compact and meaningful representation of audio data, enabling the realization of high compression ratios while retaining the integrity of the audio signal. The mathematical framework underpinning the VAE ensures that the compression process is both effective and principled, making it a promising approach for lossless audio compression in various applications. The training phase involves iterative updates to the VAE's parameters to minimize the total loss, while the compression and decompression phases utilize the trained VAE to achieve lossless compression of audio signals. **Fig. 2** illustrates the proposed algorithm of the research.

4. DEEP LEARNING PARAMETERS FOR THE RESEARCH

These are the variables that have the ability for learning, which resolve the Neural Networks (NNs) performances and behaviors. The following are the Deep Learning parameters **(Al-Bayati et al., 2020; Ghadi and Salman, 2022; Alfarhany and Abdullah, 2023; Hassan and Dawood, 2024; Yasir and Al-Barrak, 2024)**:

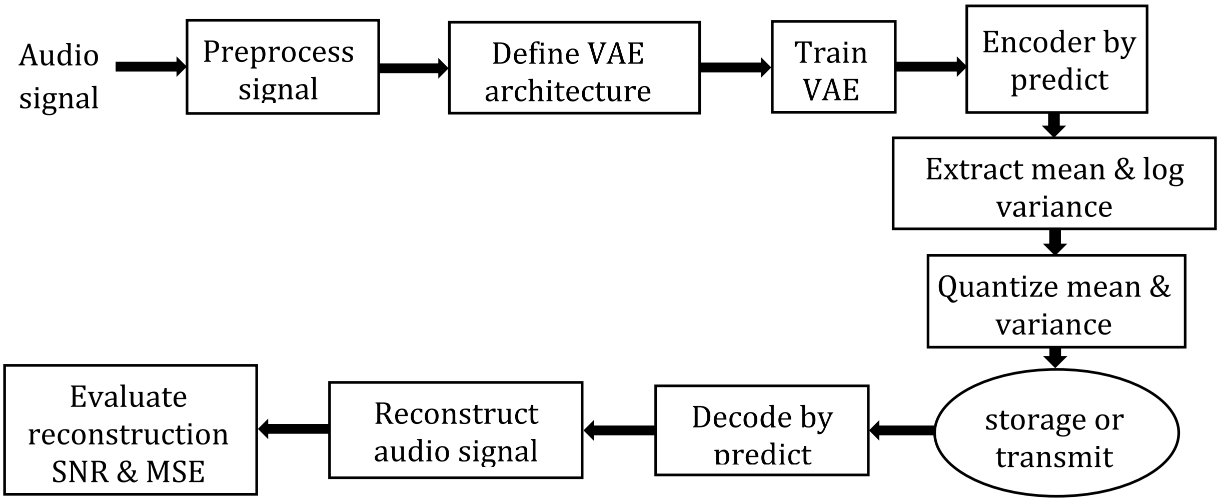

Figure 2. The research Algorithm.

4.1 Initial Observations

For the high initial loss, the initial loss value is 6.1e-03, which is relatively high compared to subsequent values. This suggests that the model starts with a significant amount of error, which is common in the early stages of training. For the rapid decrease in loss, the loss quickly decreases to values in the range of 10^{-5} to10^{-6}. This rapid decline indicates that the model is learning efficiently from the training data and that the chosen learning rate and optimization algorithm are effective.

4.2 RMSE Analysis

The RMSE values provided a minimum value of 1µ. The maximum value is 0.001. A lower RMSE is desirable as it indicates that the model's predictions are closer to the actual values.

4.3 Consistency in RMSE

The RMSE values show a consistent decrease, which is a positive sign. However, fluctuations in RMSE (e.g., from 0.0038 to 0.0039) suggest that while the overall trend is downward. There may be some variabilities in the model's performance from one epoch to the next in Model Performance.

4.4 Learning Rate

The rapid decrease in loss and RMSE suggests that the learning rate is appropriately set. If the learning rate were too high, the model might overshoot the optimal solution, leading to instability or divergence. If it were too low, the model would learn too slowly or get stuck in a suboptimal solution.

4.5 Overfitting Concerns

The relatively low RMSE values towards the end of training could indicate that the model is fitting the training data well. However, it's important to monitor for overfitting. The monitoring is especially important if the model's performance on unseen data (validation or test set) is not as good as on the training set.

4.6 Comparison with Loss

For the correlation between loss and RMSE, they generally decrease together, which is expected since both are measures of error. The specific values and the rate of decrease can vary depending on the nature of the data. The model complexity also has specific values.

5. EXPERIMENTS AND RESULTS

Our proposal has been tested using Matlab IDE, Notepad++, and Audacity DSP audio player. The tests included 119 audio files from the standard BBC audio library. The total duration of these standard audio files is about 1000 seconds (more than 150 minutes). The audio files are single-channel mono files, and the multi-channel stereo files were mixed to produce mono format (double-channel). The sampling rate is a standard 44100 sample/second with a depth of 16 bits/sample. The files were different types of audio such as bubble sound, music piano, the market in a specific country, domestic hens, flamingo birds, classic symphonies, different machines and vehicles, etc. The length of the files ranged from 0.142 to 297.4 seconds. The subjective tests for the compressed and decompressed audio denote that the method is very good. The testers were different people of different ages, cultures, and genders. Their response was it is difficult to recognize the original, compressed and recovered audio. **Fig. 3** displays the original input waveforms of a typical audio signal and the reconstructed output signal. The samples of the two waveforms are identical for all that tested typical audio signals. For the objective tests, the Signal Noise Ratio of the recovered samples was very high (from 8.04 to 50.08 dB) with a 26.33 dB average value. The mean Square Error was negligible (from -65.08 to 40.01 dB) with a -52.58-dB average value. The researchers checked 3 values of compression ratios: 8, 4, 2, and 1 (no compression). After different compressions and cross-checking, the researchers chose the highest ratio of 8. The choice is a compromising solution based on: The best SNR of the recovered audio signal, the best MSE, and the less required time for the real-time compression implementation. The time range is from 32 to 212 sec, with a 100 sec average value for each 1 sec of the input signal. To reduce that time, the Epochs were 30, hidden layers were 64 to 128, the Quantization level was unity, and Latent Dimensions were 15 to 20. The above maximum, average, and maximum values are tabulated in **Table 1**, and illustrated in **Fig. 4**.

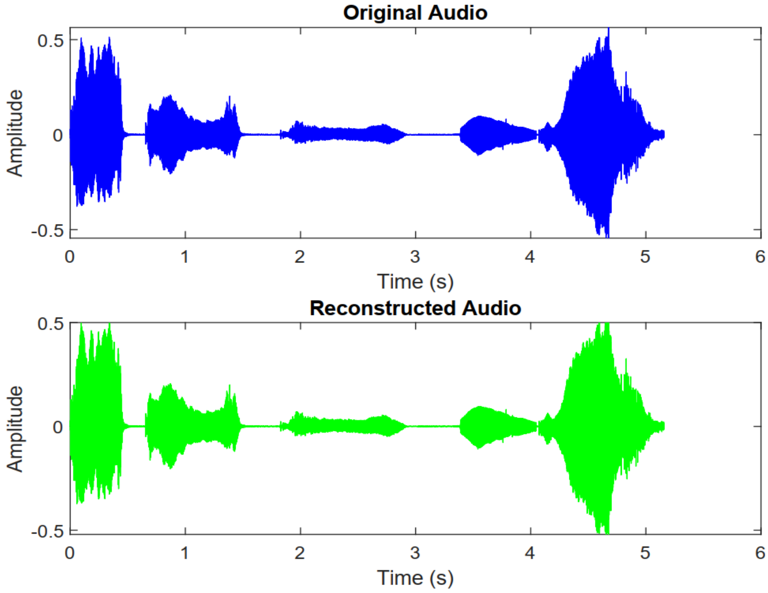

Figure 3. The original (blue) and the recovered (green) audio signals.

Table 1. Minimum, Average, and Maximum value of the Compression process for the research.

Parameter	Minimum	Average	Maximum
Duration of audio file (sec)	0.142	8.973	297.4
Signal to Noise Ratio (SNR) (dB)	8.04	26.33	50.08
Mean Square Error (MSE) (dB)	-65.08	-52.58	-40.01
Processing time (sec) for 1 sec input signal	32	100	212

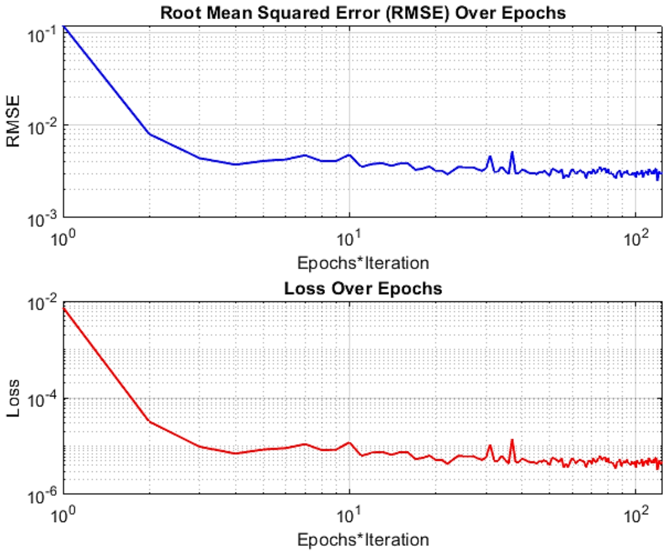

Figure 4. RMSE (blue) and Loss (red) versus the Epochs iterations.

6. CONCLUSIONS

According to compression between this paper's research with other reliable and standard research ,The method is efficient for the compression ratio, SNR, MSE, and stability against different types of audio. The maximum obtained CR is 8, with 2×10^{-3} RMSE and 5×10^{-6} loss over 30 epochs, 128 hidden layers, and 20 latent dimensions. The weak point of this method is the long-required time for real-time implementation. The compression ratio is acceptable compared with other famous algorithms and techniques. The loss and the mean square error of the reconstructed audio signal could be omitted compared with the original input audio signal. The tables, figures, and waveforms denote successful and stable performance, versus the other lossless audio compression systems and techniques.

NOMENCLATURE

Symbol	Description	Symbol	Description
$diag$	diagonal matrix.	X, $\hat{X}$	Input, and estimated input signals.
L_{total}	Total loss.	Z	Latent representation.
L_{rec} & L_{KL}	Reconstruction & Kullback-Leibler divergence losses.	λ	Hyperparameter to control the trade-off between the components.
M	Number of samples per batch.	σ	Variance.
N	Normal distribution.	μ	Mean value.

Acknowledgements

This work was supported by the (Electrical Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq).

Credit Authorship Contribution Statement

Ali A. Obaid: Writing – review & editing, Writing – original draft, Validation, Software, Methodology. Hasan M. Kadhim: Writing – review & editing, Writing – final draft, Validation, Software, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Al-Bayati, A.Q., Al-Araji, A.S. and Ameen, S.H., 2020. Arabic sentiment analysis (ASA) using a deep learning approach. *Journal of Engineering*, *26*(6), pp.85-93. https://doi.org/10.31026/j.eng.2020.06.07.

Alfarhany, A.A.R. and Abdullah, N.A., 2023. Iraqi sentiment and emotion analysis using deep learning. *Journal of Engineering*, *29*(09), pp.150-165. https://doi.org/10.31026/j.eng.2023.09.11.

Amada, S., Sugiura, R., Kamamoto, Y., Harada, N., Moriya, T., Yamada, T. and Makino, S., 2018. Experimental evaluation of waiver predictor for audio lossless coding. In *The Acoustical Society of Japan 1018Autumn Meeting*, pp. 1149-1152.

Barman, R., Badade, S., Deshpande, S., Agarwal, S. and Kulkarni, N., 2022. Lossless data compression method using deep learning. In *Machine Intelligence and Smart Systems: Proceedings of MISS 2021* (pp. 145-151). Singapore: Springer Nature Singapore. http://dx.doi.org/10.1007/978-981-16-9650-3_11.

Chen, Q., Wu, W. and Luo, W., 2021. Lossless compression of sensor signals using an untrained multi-channel recurrent neural predictor. *Applied Sciences*, *11*(21), p.10240. https://doi.org/10.3390/app112110240.

Crocco, M., Cristani, M., Trucco, A. and Murino, V., 2016. Audio surveillance: A systematic review. *ACM Computing Surveys (CSUR)*, *48*(4), pp.1-46. https://doi.org/10.1145/2871183.

Cunha, B.Z., Droz, C., Zine, A.M., Foulard, S. and Ichchou, M., 2023. A review of machine learning methods applied to structural dynamics and vibroacoustic. *Mechanical Systems and Signal Processing*, *200*, p.110535. http://dx.doi.org/10.1016/j.ymssp.2023.110535.

Défossez, A., Copet, J., Synnaeve, G. and Adi, Y., 2022. High-fidelity neural audio compression. *arXiv preprint arXiv:2210.13438*. https://doi.org/10.48550/arXiv.2210.13438.

Dewangan, G. and Maurya, S., 2021. Fault diagnosis of machines using deep convolutional beta-variational autoencoder. *IEEE Transactions on Artificial Intelligence*, *3*(2), pp.287-296. https://doi.org/10.1109/TAI.2021.3110835

Dubois, Y., Bloem-Reddy, B., Ullrich, K. and Maddison, C.J., 2021. Lossy compression for lossless prediction. *Advances in Neural Information Processing Systems*, *34*, pp.14014-14028. https://doi.org/10.48550/arXiv.2106.10800.

Friedland, G., Jia, R., Wang, J., Li, B. and Mundhenk, N., 2020, August. On the impact of perceptual compression on deep learning. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 219-224). IEEE. https://doi.org/10.1109/MIPR49039.2020.00052.

Ghadi, N.M. and Salman, N.H., 2022. Deep learning-based segmentation and classification techniques for brain tumor MRI: A review. *Journal of Engineering*, *28*(12), pp.93-112. https://doi.org/10.31026/j.eng.2022.12.07.

Hassan, B.A.R. and Dawood, F.A.A., 2024. Face-based gender classification using deep learning model. *Journal of Engineering*, *30*(01), pp.106-123. https://doi.org/10.31026/j.eng.2024.01.07.

Hemmer, M., Klausen, A., Van Khang, H., Robbersmyr, K.G. and Waag, T.I., 2020. Health indicator for low-speed axial bearings using variational autoencoders. *IEEE Access*, *8*, pp.35842-35852. https://doi.org/10.1109/ACCESS.2020.2974942.

Hennequin, R., Royo-Letelier, J. and Moussallam, M., 2017, March. Codec-independent lossy audio compression detection. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 726-730). IEEE. https://doi.org/10.1109/ICASSP.2017.7952251.

Huang, Q., Liu, T., Wu, X., and Qu, T., 2019. A generative adversarial net-based bandwidth extension method for audio compression. *Journal of the Audio Engineering Society*, *67*(12), pp.986-993. https://doi.org/10.17743/jaes.2019.0047.

Jain, A. and Patel, R., 2009, May. An efficient compression algorithm (ECA) for text data. In *2009 international conference on signal processing systems* (pp. 762-765). IEEE. https://doi.org/10.1109/ICSPS.2009.96.

Jing, W., Xiang, X. and Jingming, K., 2014. A novel multichannel audio signal compression method based on tensor representation and decomposition. *China Communications*, *11*(3), pp.80-90. https://doi.org/10.1109/CC.2014.6825261.

Kalinin, S.V., Dyck, O., Jesse, S. and Ziatdinov, M., 2021. Exploring order parameters and dynamic processes in disordered systems via variational autoencoders. *Science Advances*, *7*(17), p.eabd5084. https://doi.org/10.1126/sciadv.abd5084

Liu, Y., 2021, November. Recovery of lossy compressed music based on CNN super-resolution and GAN. In *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)* (pp. 623-629). IEEE. https://doi.org/10.1109/ICFTIC54370.2021.9647041.

Nagaraj, P., Rao, J.S., Muneeswaran, V. and Kumar, A.S., 2020, May. Competent ultra data compression by enhanced features excerption using deep learning techniques. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1061-1066). IEEE. https://doi.org/10.1109/ICICCS48265.2020.9121126.

Nogales, A., Donaher, S. and García-Tejedor, Á., 2023. A deep learning framework for audio restoration using Convolutional/Deconvolutional Deep Autoencoders. *Expert Systems with Applications*, *230*, p.120586. https://doi.org/10.1016/j.eswa.2023.120586.

Passricha, V. and Aggarwal, R.K., 2019. A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition. *Journal of Intelligent Systems*, *29*(1), pp.1261-1274. https://doi.org/10.1515/jisys-2018-0372.

Pollastro, A., Testa, G., Bilotta, A. and Prevete, R., 2023. Semi-supervised detection of structural damage using variational autoencoder and a one-class support vector machine. *IEEE Access*. https://doi.org/10.48550/arXiv.2210.05674.

Ramesh, V.; Wang, M., 2021. Recurrent autoencoders with dynamic time warping for near-lossless music compression and minimal-latency transmission. Preprints 2021, ClefNet, 2021030360. https://doi.org/10.20944/preprints202103.0360.v1

San Martin, G., López Droguett, E., Meruane, V. and das Chagas Moura, M., 2019. Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring*, *18*(4), pp.1092-1128. https://doi.org/10.1177/1475921718788299.

Schuller, G.D., Yu, B., Huang, D. and Edler, B., 2002. Perceptual audio coding using adaptive pre-and post-filters and lossless compression. *IEEE Transactions on Speech and Audio Processing*, *10*(6), pp.379-390. https://doi.org/10.1109/TSA.2002.803444.

Shang, Z., Sun, L., Xia, Y., and Zhang, W., 2021. Vibration-based damage detection for bridges by deep convolutional denoising autoencoder. *Structural Health Monitoring*, *20*(4), pp.1880-1903. https://doi.org/10.1177/1475921720942836.

Shin, S., Byun, J., Park, Y., Sung, J. and Beack, S., 2022, May. Deep neural network (DNN) audio coder using a perceptually improved training method. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 871-875). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9747575.

Shukla, S., Ahirwar, M., Gupta, R., Jain, S. and Rajput, D.S., 2019, February. Audio compression algorithm using discrete cosine transform (DCT) and Lempel-Ziv-Welch (LZW) encoding method. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 476-480). IEEE. https://doi.org/10.1109/COMITCon.2019.8862228.

Shukla, S., Gupta, R., Rajput, D.S., Goswami, Y. and Sharma, V., 2022. A comparative analysis of lossless compression algorithms on uniformly quantized audio signals. *International Journal of Image, Graphics and Signal Processing*, *13*(6), p.59. https://doi.org/10.5815/ijigsp.2022.06.05.

Välimäki, V. and Reiss, J.D., 2016. All about audio equalization: Solutions and frontiers. *Applied Sciences*, *6*(5), p.129. https://doi.org/10.3390/app6050129.

Yasir, M.H. and Al-Barrak, A., 2024. Utilizing deep learning techniques to identify people by palm print. *Journal of Engineering*, *30*(04), pp.87-98. https://doi.org/10.31026/j.eng.2024.04.06.

Yoshimura, T., Hashimoto, K., Oura, K., Nankaku, Y. and Tokuda, K., 2018, December. WaveNet-based zero-delay lossless speech coding. In *2018 IEEE Spoken Language Technology Workshop (SLT)* (pp. 153-158). IEEE. https://doi.org/10.1109/SLT.2018.8639598.

Zeghidour, N., Luebs, A., Omran, A., Skoglund, J. and Tagliasacchi, M., 2021. Soundstream: An end-to-end neural audio codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, pp.495-507. https://doi.org/10.1109/TASLP.2021.3129994.

التعلم العميق لضغط الصوت بدون فقدان

علي أحمد عبيد*، حسن محمدعلي كاظم

قسم الهندسة الكهربائية، كلية الهندسة، الجامعة المستنصرية، بغداد، العراق

الخلاصة

يتم استخدام تقنيات ضغط الصوت والكلام لتقليل تخزين هذه البيانات في المساحة المطلوبة وتقليل معدل نقل هذه البيانات في أنظمة الاتصالات والشبكات. في هذا البحث، يستغل الباحثون الشبكات العصبية والذكاء الاصطناعي لضغط الإشارات الصوتية. قام الباحثون بدراسة نسبة الضغط 8، 4، 2 و1 (بدون ضغط)، ومن ثم اختيار أعلى نسبة 8. يعتمد اختيار التسوية على أفضل نسبة إشارة صوتية مستردة الى الضوضاء SNR، والوقت المطلوب للتنفيذ. اختبر الباحثون 119 ملفًا صوتيًا مختلفًا من مكتبة بي بي سي الصوتية القياسية. مدة هذه الملفات حوالي 1000 ثانية. متوسط: نسبة SNR كانت 26.33 ديسيبل، ومتوسط الخطأ المربع هو –52.6 ديسيبل. لتقليل وقت التشغيل، كانت العصور 30، والطبقات المخفية من 64 إلى 128، ومستوى التكميم هو 1، والأبعاد الكامنة من 15 إلى 20، وكل ثانية واحدة من إشارة الدخل تحتاج إلى معدل 211 ثانية ليتم ضغطها وفك ضغطها. كانت ملفات الإشارة الصوتية المدخلة صوت أحادي القناة وتمت إعادة تنسيق ملفات الصوت الستريو متعددة القنوات إلى قناة أحادية مونو. نتائج البحث حققت ضغطاً صوتياً جيداً بينما كانت المعلمات الأخرى مقبولة.

الكلمات المفتاحية: التعلم العميق، ضغط الصوت، أجهزة الترميز التلقائي المتغيرة، نسبة الضغط.

APPENDIX

Arbitrary samples of the tested audio files. Duration of each audio, MSE of the compression (×10^{-3}), SNR of the recovered decompressed audio, and the processing time for each second. The Hidden Layers = 128, the Latent Dimensions = 20, the Epochs = 30, and the Compression Ratio = 8. The researchers exploited the BBC audio dataset to test their proposed algorithm. The library contains different high-quality audio.

File name (.wav)	Duration/S	MSE	SNR/Db	Time/s
Stage Show.wav	2.190	0.056	8.14	628
Music Hassjk.wav	2.700	0.006	25.32	822
Flamingo.wav	0.998	0.013	19.25	1580
First Aid.wav	2.440	0.001	47.52	676
Bubbles.wav	0.142	0.003	37.52	40
Big-Ben.wav	4.420	0.005	27.20	1658
09 Bells of Sant Gervais Church.wav	2.000	0.007	24.54	851
Automatic Washing Machine.wav	1.020	0.003	32.72	274
Amusement Arcade (Sitia) .wav	2.430	0.022	14.76	727
Man Speech .wav	1.490	0.006	29.11	96
General Atmosphere	1.100	0.007	41.98	411
Wood File.wav	4.500	0.013	18.52	1714
Biano.wav	1.990	0.004	38.84	778
Day Old Boy (Restless).wav	4.900	0.019	16.92	798
Week Old Girl (Hiccoughs).wav	2.640	0.004	33.17	334
Month Old Girl, Talking Nonsense.wav	3.990	0.006	22.87	627
1-Year Old Boy (Laughing).wav	8.026	0.019	16.70	2782
Heart Beat (Average).wav	1.470	0.029	15.50	592
Fetal Heart Beat.wav	0.502	0.011	20.56	71
Men's Ward.wav	4.000	0.011	22.55	1423
Morocco (Medina Food).wav	1.090	0.029	33.21	401
Morocco Street (Menkes).wav	1.000	0.003	31.80	319
Algeria (Market).wav	1.700	0.020	15.69	604
Zaire (Small But).wav	2.290	0.004	28.63	638
Cameroun (Insects and Bard).wav	1.002	0.007	23.23	186
Senegal (Insects and Bard).wav	6.520	0.003	27.80	2006
Elephants (Elephants Clos).wav	7.590	0.004	29.27	2591
Burchell's Zebra (Close Up Calls).wav	2.000	0.002	33.97	606
Chimpanzees (Close Up).wav	4.600	0.001	33.38	1765