JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

A COMPARATIVE ANALYSIS OF OPTIMIZED NEURAL NETWORK AND LARGE-SCALE LANGUAGE MODELS FOR MUSIC GENRE CLASSIFICATION

Ahmad Naufal Luthfan Marzuqi[1], Vinna Rahmayanti Setyaning Nastiti [1*]
[1]Informatics Engineering, Universitas Muhammadiyah Malang
email: *vinastiti@umm.ac.id

Abstract: The rapid growth of the digital music industry requires accurate music genre classification systems to enhance user experience in streaming services. This study compares a domain-specific Long Short-Term Memory (LSTM) network with three Large Language Models (LLMs)—HuBERT, WavLM, and WAV2Vec 2.0—for Music Genre Classification (MGC). The LSTM model was trained using Mel-spectrograms transformed from the GTZAN dataset, while the LLMs were fine-tuned using a smaller set of raw audio samples due to computational constraints. All models were tested on datasets with identical genre labels to ensure a fair evaluation. Results show that the LSTM model achieved the highest accuracy of **97.10%**, outperforming HuBERT (86.00%), WavLM (83.00%), and WAV2Vec 2.0 (80.00%). The LSTM demonstrated superior generalization and stability without overfitting, while the LLMs struggled to differentiate between genres with similar acoustic characteristics. These findings indicate that general-purpose pre-trained models, although powerful, are less effective in music-specific tasks due to domain mismatch. Therefore, incorporating music-specific features and architectures remains essential for achieving higher accuracy and reliability in automatic genre classification systems.

Keywords: audio large language models; comparative deep learning; music genre classification.

Abstrak: Pertumbuhan industri musik digital yang pesat menuntut sistem klasifikasi genre musik yang akurat untuk meningkatkan pengalaman pengguna dalam layanan streaming. Penelitian ini dilatarbelakangi oleh perkembangan pesat model pembelajaran mendalam, khususnya jaringan LSTM dan model bahasa berskala besar LLM seperti HuBERT, WavLM, dan WAV2Vec 2.0, yang telah menunjukkan kemampuan representasi audio yang kuat. Tujuan penelitian ini ini membandingkan jaringan Long Short-Term Memory (LSTM) khusus domain dengan tiga model Large Language Models (LLM)—HuBERT, WavLM, dan WAV2Vec 2.0—untuk tugas Klasifikasi Genre Musik (MGC). Metode penelitian melibatkan pelatihan LSTM menggunakan data Mel-spectrogram hasil transformasi dari dataset GTZAN, sementara LLM disesuaikan (fine-tuning) menggunakan data audio mentah dalam jumlah lebih kecil karena keterbatasan komputasi. Seluruh model diuji pada dataset dengan label genre yang sama untuk memastikan evaluasi yang adil. Hasil penelitian menunjukkan bahwa model LSTM mencapai akurasi tertinggi sebesar 97,10%, sedangkan model HuBERT, WavLM, dan WAV2Vec 2.0 masing-masing memperoleh 86,00%, 83,00%, dan 80,00%. Model LSTM menunjukkan kemampuan generalisasi yang lebih baik tanpa overfitting, sedangkan model LLM cenderung kesulitan membedakan genre dengan karakteristik akustik yang mirip. Kesimpulan penelitian ini adalah ketidaksesuaian domain secara signifikan membatasi performa model umum saat diterapkan pada tugas berbasis musik. Oleh karena itu, penggunaan fitur dan arsitektur khusus musik sangat penting dalam membangun sistem klasifikasi genre yang lebih akurat.

Kata kunci: klasifikasi genre musik; model bahasa besar; perbandingan pembelajaran mendalam.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The global recorded music market increased by 10.2% in 2023, surpassing 700 million paid streaming subscribers [18]. This underscores the significance of Music Genre Classification (MGC) within Music Information Retrieval (MIR), enhancing user experience through better recommendations and library organization [15]. MGC has progressed from traditional techniques like Mel-Frequency Cepstral Coefficients (MFCCs) and Support Vector Machines (SVMs) [7] to advanced deep learning models such as Convolutional Neural Networks (CNNs) [5] and Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM), which effectively recognize temporal patterns in music.

Recently, self-supervised audio foundation models, commonly known as Audio Large Language Models (LLMs), have been developed, such as HuBERT, WavLM, and Wav2Vec 2.0 [16]. These models were chosen for this research because they exemplify different approaches in self-supervised learning: HuBERT uses masked prediction, WavLM integrates denoising and speaker-aware training, and Wav2Vec 2.0 relies on contrastive learning.

This study compares an optimized LSTM-based model with three fine-tuned audio LLMs for MGC, providing insights into the balance between specialization and generalization in designing more effective MGC systems.

METHOD

This study compares two methods: a Neural Network (LSTM) model and a Large Language Model (LLM). The LSTM analyzes temporal patterns in audio signals, whereas the LLM is adapted to understand and categorize textual audio features.

The main dataset is GTZAN, comprising 1,000 thirty-second clips spanning 10 genres. Its balanced distribution, illustrated in Image 1, aids in preventing genre bias.

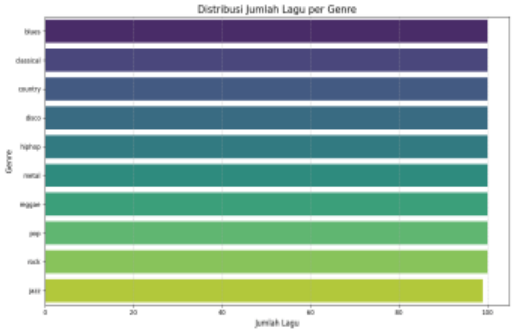

Image 1. Distribution of Songs per Genre

Audio can take different forms. Image 2 displays the raw waveform showing amplitude variations over time. For the LSTM model, these waveforms were transformed into Mel-Spectrograms.

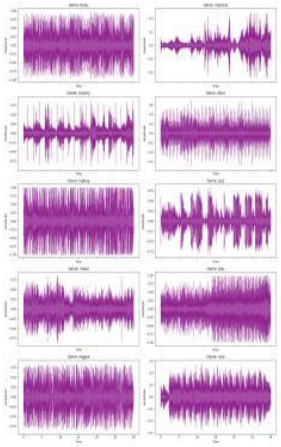

Image 2. Visualization of Audio Waveforms for Each Genre

A Mel-spectrogram Image 3 shows how frequency changes over time, making it suitable for deep learning models that process images or sequential data.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

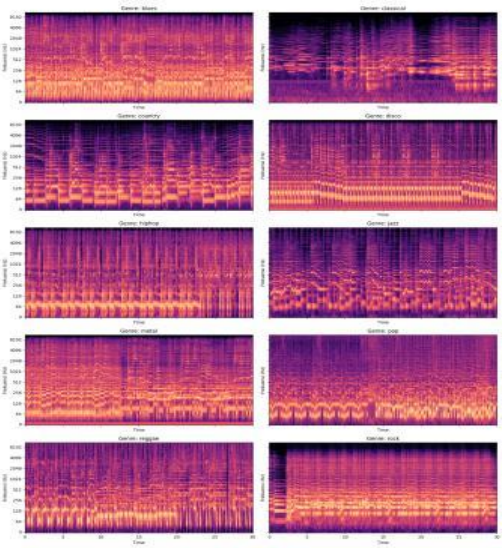

Image 3. Visualization of Mel-Spectrograms for Each Genre

Each pipeline processed data differently: the LSTM model was trained on an augmented dataset of approximately 5,000 segmented samples, while the LLMs underwent fine-tuning on a smaller subset of 200 raw audio samples because of their higher computational requirements.

RESULT AND DISCUSSION

The custom-trained NN (LSTM) model performed well, converging after. 52 epochs. Image 4 shows stable validation accuracy and loss, indicating good generalization without overfitting.

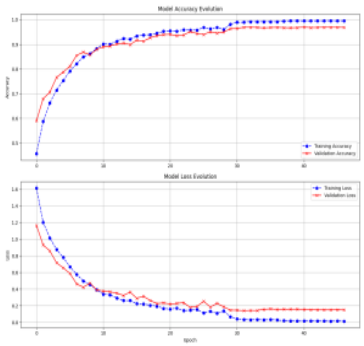

Image 4. Accuracy and Loss Evolution of the NN (LSTM) Model

The LSTM model showed successful learning and generalization without overfitting, indicated by the convergence of training and validation curves. It reached a test accuracy of 97.10%, as evidenced by the prominent diagonal in the confusion matrix (Image 5).

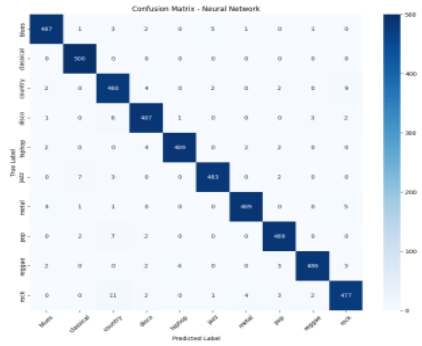

Image 5. Confusion Matrix of the Neural Network (LSTM)

This confusion matrix shows the LSTM model's classification, with 97.10% accuracy. Correct predictions, marked by dark blue diagonal cells, indicate few misclassifications, proving the model's effectiveness in distinguishing music genres.

The data in Table 1's classification report confirms these results, highlighting high F1-scores for all classes.

Table 1. Classification Neural Network

Genres	Table Column Title		
	Precision	Recall	F1-Score
Blues	0.99	0.96	0.97
Classical	0.99	0.99	0.99
Country	0.94	0.96	0.95
Disco	0.97	0.95	0.96
Hiphop	0.99	0.98	0.98
Jazz	0.97	0.97	0.97
Metal	1.00	0.98	0.99
Pop	0.97	0.98	0.98
Reggae	0.98	0.96	0.97
Rock	0.94	0.96	0.95
Accuracy			0.97

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

Macro avg	0.97	0.97	0.97
Weighted avg	0.97	0.97	0.97

The table shows the LSTM model's excellent performance across genres, with F1-scores over 0.95 and an overall accuracy of 0.97, making it the most precise model.

Fine-tuned LLMs like HuBERT, WavLM, and WAV2Vec 2.0 are robust but less precise than LSTM models. LSTMs are more accurate but require more preprocessing and training; LLMs process raw audio faster and cheaper. This shows the trade-off between accuracy and resources.

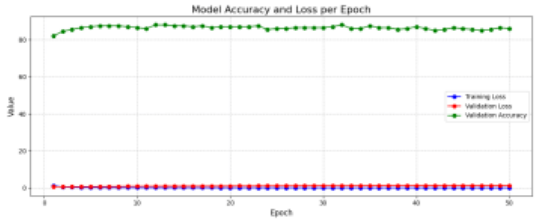

Image 6. Training and Validation Curves for HuBERT

The training graph shows fine-tuning's benefit, quickly reaching 86% validation accuracy. Flat, low-loss curves indicate pre-trained features fit the task well. Minimal training-validation difference confirms no overfitting.

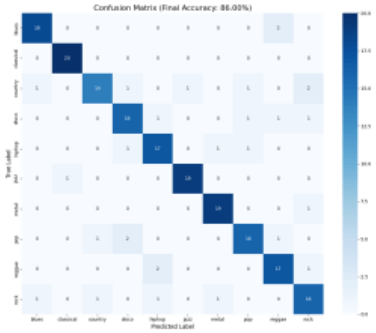

Image 7. Confusion Matrix for HuBERT

The confusion matrix shows an 86.00% accuracy, highlighting the model's strength in classifying clear genres like classical, metal, and jazz. However, it struggled with similar genres like country, rock, and blues, causing most misclassifications.

Table 2. Classification HuBERT

Genres	Table Column Title		
	Precision	Recall	F1-Score
Blues	0.90	0.90	0.90
Classical	0.95	1.00	0.98
Country	0.88	0.70	0.78
Disco	0.80	0.80	0.80
Hiphop	0.81	0.85	0.83
Jazz	0.95	0.95	0.95
Metal	0.90	0.95	0.93
Pop	0.84	0.80	0.82
Reggae	0.81	0.85	0.83
Rock	0.76	0.80	0.78
Accuracy			0.86
Macro avg	0.86	0.86	0.86
Weighted avg	0.86	0.86	0.86

HuBERT achieved 86% accuracy but was inconsistent, especially with lower F1-scores in country, disco, pop, and rock.

WavLM, based on HuBERT and pre-trained on noisy audio with denoising, was fine-tuned for music classification but didn't match specialized neural network performance.

Image 8. Training and Validation Curves for WavLM

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

The WavLM training graph shows rapid early learning but also indicates overfitting. Validation accuracy quickly hit a fluctuating peak and then stabilized at 84%, while the training loss steadily declined. Meanwhile, the validation loss continued to increase after the first few epochs, displaying a typical overfitting trend.

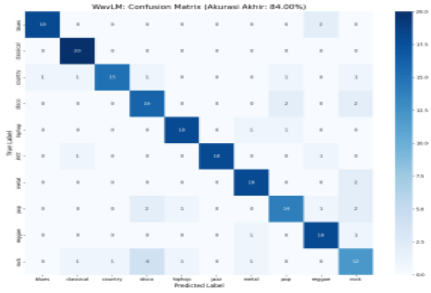

Image 9. Confusion Matrix for WavLM Model

The confusion matrix confirms an 84.00% accuracy, indicating WavLM performed well on genres such as classical, blues, hip-hop, and metal. Its primary weakness was differentiating similar genres, particularly confusing 'rock' with 'disco' and 'pop'.

Table 3. Classification WavLM

Genres	Table Column Title		
	Precision	Recall	F1-Score
Blues	0.90	0.90	0.90
Classical	0.95	1.00	0.98
Country	0.88	0.70	0.78
Disco	0.80	0.80	0.80
Hiphop	0.81	0.85	0.83
Jazz	0.95	0.95	0.95
Metal	0.90	0.95	0.93
Pop	0.84	0.80	0.82
Reggae	0.81	0.85	0.83
Rock	0.76	0.80	0.78
Accuracy			0.86
Macro avg	0.86	0.86	0.86
Weighted avg	0.86	0.86	0.86

WavLM reached 84% accuracy, which is lower than HuBERT, and showed decreased F1-scores in 'disco', 'pop', and 'rock'. This highlights the limited capacity of generalist LLMs to grasp musical subtleties compared to specialized neural networks.

Wav2Vec 2.0, a self-supervised model fine-tuned for this task, underperformed compared to the specialized neural network in terms of accuracy and genre-level consistency.

Image 10. Training and Validation Curves for WAV2Vec 2.0

The wav2vec 2.0 training graph shows rapid early gains, then plateaus at 80.00%. Validation loss rises after epoch 15, indicating overfitting as the model memorizes rather than generalizes.

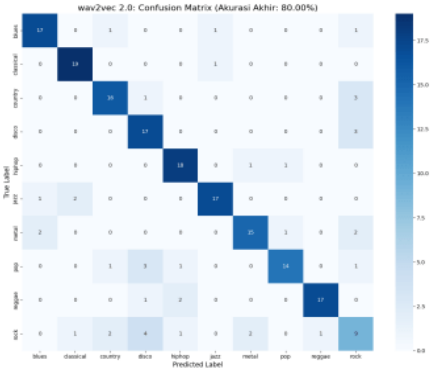

Image 11. Confusion Matrix for WAV2Vec 2.0 Model

The confusion matrix shows 80% accuracy, reflecting strong performance on classical and hip-hop genres. However, it struggles to differentiate similar styles, often confusing 'rock' with 'country' and

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XI No 4, September 2025, hlm. 741 – 748 ISSN 2550-0201 (Online)
DOI: https://doi.org/ 10.33330/jurteksi.v11i4.4198
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

'metal,' and misclassifies 'pop' as 'disco' due to acoustic similarities.

Table 4. Classification Wav2vec 2.0

Genres	Table Column Title		
	Precision	Recall	F1-Score
Blues	0.85	0.85	0.85
Classical	0.86	0.95	0.90
Country	0.80	0.80	0.80
Disco	0.65	0.85	0.74
Hiphop	0.82	0.90	0.86
Jazz	0.89	0.85	0.87
Metal	0.83	0.75	0.79
Pop	0.88	0.70	0.78
Reggae	0.94	0.85	0.89
Rock	0.47	0.45	0.46
Accuracy			0.80
Macro avg	0.80	0.80	0.79
Weighted avg	0.80	0.80	0.79

WAV2Vec 2.0 achieved 80% accuracy but underperformed, particularly in 'rock' (F1 0.46), 'disco', and 'pop', exposing its limitations in music classification.

Table 5. Final Accuracy of All Models

Model	Accuracy
NN (LSTM)	97,10%
HuBERT	86.00%
WavLM	83.00%
WAV2Vec	80.00%

The comparison highlights the specialized LSTM model's top performance with 97.10% accuracy, exceeding all LLMs: HuBERT (86.00%), WavLM (83.00%), and WAV2Vec 2.0 (80.00%). It improves by +6.84% over the 90.26% reported in [7]. WavLM's 83.00% is slightly below by -1.6% the 84.6% noted in [16].

CONCLUSION

This research compared a specialized LSTM model with three fine-tuned large language models (LLMs) for music genre classification. The LSTM significantly outperformed the best LLM, achieving 97.10% accuracy compared to 86.00%. This indicates that, although general-purpose LLMs are powerful, using domain-specific data is essential for optimal results, as speech-focused pre-training restricts effectiveness in the music domain.

BIBLIOGRAPHY

[1] M. Wu and X. Liu, “A Double Weighted KNN Algorithm and Its Application in the Music Genre Classification,” in 2019 6th International Conference on Dependable Systems and Their Applications (DSA), Harbin, China: IEEE, Jan. 2020, pp. 335–340. doi: 10.1109/DSA.2019.00051.

[2] N. Narkhede, S. Mathur, A. A. Bhaskar, K. K. Hiran, M. Dadhich, and M. Kalla, “A New Methodical Perspective for Classification and Recognition of Music Genre Using Machine Learning Classifiers,” in 2023 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia: IEEE, Aug. 2023, pp. 94–99. doi: 10.1109/ETNCC59188.2023.10284969.

[3] U. M. Srinivas, S. Rafi, T. V. Manohar, and M. V. Rao, “Classification of Music Genre Using Deep Learning Approaches,” in 2024 4th International

Conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India: IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/AISP61711.2024.10870721.

[4] I. Pathania and N. Kaur, "Classification of Music Genre Using Machine Learning," in 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India: IEEE, Oct. 2022, pp. 1–5. doi: 10.1109/GCAT55367.2022.9972105.

[5] M. Singla, K. S. Gill, M. Kumar, and R. Rawat, "Classification of Musical Genres Utilizing the CNN Sequential Model and Deep Learning Techniques," in 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India: IEEE, Jun. 2024, pp. 1–5. doi: 10.1109/ICITEICS61368.2024.10625371.

[6] Z. Ma, "Comparison between Machine Learning Models and Neural Networks on Music Genre Classification," in 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China: IEEE, May 2022, pp. 189–194. doi: 10.1109/CVIDLICCEA56201.2022.9825050.

[7] R. Gusain, S. Sonker, S. K. Rai, A. Arora, and S. T. Nagarajan, "Comparison of Neural Networks and XGBoost Algorithm for Music Genre Classification," in 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India: IEEE, Jun. 2022, pp. 1–6. doi: 10.1109/CONIT55038.2022.9847814.

[8] S. Mohanapriya, S. Jhansi Ida, M. Magadalene, S. Nithiyashree, U. Monisha, and M. Indraja, "Deep Learning-Based Music Genre Classification using Convolutional Neural Network," in 2024 First International Conference on Software, Systems and Information Technology (SSITCON), Tumkur, India: IEEE, Oct. 2024, pp. 1–6. doi: 10.1109/SSITCON62437.2024.10796579.

[9] V. Shah, A. Tandle, N. Sharma, and V. Sheth, "Genre Based Music Classification using Machine Learning and Convolutional Neural Networks," in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, Jul. 2021, pp. 1–8. doi: 10.1109/ICCCNT51525.2021.9579597.

[10] M. Sambath, R. L. Kumar, S. M. Vishnu Reddy, V. P. Reddy, L. Joseph, and M. Kathiravan, "Identification and Classification of Music Genre using Deep Learning," in 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India: IEEE, Sep. 2022, pp. 1–6. doi: 10.1109/ICCSEA54677.2022.9936530.

[11] N. Srivastava, S. Ruhil, and G.

Kaushal, "Music Genre Classification using Convolutional Recurrent Neural Networks," in 2022 IEEE 6th Conference on Information and Communication Technology (CICT), Gwalior, India: IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/CICT56698.2022.9997961.

[12] A. Ghildiyal and S. Sharma, "Music Genre Classification Using Data Filtering Algorithm: An Artificial Intelligence Approach," in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India: IEEE, Sep. 2021, pp. 922–926. doi: 10.1109/ICIRCA51532.2021.9544592.

[13] K. S. Mounika, S. Deyaradevi, K. Swetha, and V. Vanitha, "Music Genre Classification Using Deep Learning," in 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India: IEEE, Oct. 2021, pp. 1–7. doi: 10.1109/ICAECA52838.2021.9675685.

[14] S. Prince, J. J. Thomas, S. J. J, K. P. Priya, and J. J. Daniel, "Music Genre Classification using Deep learning - A review," in 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India: IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/CSITSS57437.2022.10026394.

[15] M. E. A. Meguenani, A. de S. Britto, and A. L. Koerich, "Music Genre Classification using Large Language Models," 2024, arXiv. doi: 10.48550/ARXIV.2410.08321.

[16] S. W. J, P. K. R M, P. K. K, and P. J, "Music Genre Classification Using LSTM and CNN," in 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India: IEEE, Jun. 2023, pp. 205–209. doi: 10.1109/ICPCSN58827.2023.00039.

[17] N. Ndou, R. Ajoodha, and A. Jadhav, "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches," in 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada: IEEE, Apr. 2021, pp. 1–6. doi: 10.1109/IEMTRONICS52119.2021.9422487.

[18] International Federation of the Phonographic Industry (IFPI), Global Music Report 2024: State of the Industry, London, UK, Mar. 2024. [Online]. Available: https://www.ifpi.org/wp-content/uploads/2024/04/GMR_2024_State_of_the_Industry.pdf