Report of Research Paper

1. Kang et al. (2024)

Kang and colleagues presented an overview of recent work in music emotion recognition. They explained how different studies use different labeling schemes, which makes the results difficult to compare. The paper also pointed out that most researchers depend on large datasets collected from the internet. Because the recordings come from many sources, their sound quality and style vary a lot. The authors suggested that controlled audio samples may give clearer insights into how musical features relate to emotions.[1]

2. Louro et al. (2024)

This work compared traditional machine-learning methods with modern deep-learning models. The authors used features such as MFCCs and spectrograms. They reported that deep networks performed better in most cases, but they required a large amount of training data. The study also mentioned that these models often behave like “black boxes,” and it is not easy to explain why they predict a certain emotion. This becomes a limitation when someone wants to study the features themselves.[2]

3. Jia et al. (2022)

Jia and co-authors built an emotion-classification system using neural networks. Their experiments were conducted on popular public datasets. While the accuracy was good, the authors also mentioned that the music samples in these datasets were produced in very different environments. Because of this variation, it was hard to understand whether the emotion came from tempo, pitch, or some other factor. The paper did not include experiments using controlled or self-made musical pieces.[3]

4. “Towards Unified Emotion Recognition Across Datasets” (2025)

This study examined whether a system trained on one dataset can work reliably on another. The authors tested several combinations and found that performance dropped heavily when switching datasets. They argued that differences in production style, recording quality, and emotional labelling make cross-dataset research difficult. This highlighted the need for more uniform or controlled data.[4]

5. Multimodal MER Survey (2025)

This survey discussed approaches that combine audio with lyrics or video to identify emotions. While multimodal systems usually perform better, they also become more complex. The authors pointed out that many papers focus on songs with lyrics. Pure instrumental emotion recognition, which depends only on audio features, is still not explored deeply.[5]

6. Deep-Learning Model for MER (2021)

This paper introduced a deep-learning pipeline using CNN and LSTM layers. The experiments showed strong results on the selected dataset. However, the authors mentioned that the model required considerable computation and long training time. The paper also did not explore how specific audio features influence the final prediction, so interpretability remained limited.[6]

7. ADFF: Deep Feature Fusion Approach (2022)

In this study, the researchers used classical machine-learning methods and clustering algorithms. They extracted features such as spectral contrast, tempo, and MFCCs. The results changed noticeably when the dataset was normalized differently. The authors stated that the variety in musical styles made clustering less stable. This again highlighted the issue of uncontrolled data.[7]

8. Music Emotion Recognition: Robust Standards Study

This paper presents a detailed discussion on the problems and future directions in music emotion recognition. The authors explain that many existing MER systems depend heavily on specific datasets and fixed taxonomies, which do not always match how listeners actually perceive emotions. They highlight that emotional responses vary from person to person and also depend on the listening context. Because of this, models trained on one dataset may not work well in real situations. The paper also points out that several studies use inconsistent labels, making it difficult to compare results across researchers. The authors suggest that emotion recognition should move toward more personalized and context-aware approaches, with clearer standards for annotation and evaluation.[8]

9. Real-Time Music Emotion Recognition using Feature Fusion (2025)

This paper presented a system that tries to recognize the emotion of a music clip in real time. The authors combined several types of audio features, such as MFCCs, energy levels, spectral patterns, and rhythm-based features. They used a Bi-LSTM model to process these features together. The system gave good accuracy on the test dataset, but the authors also mentioned that the approach was quite complex. It required fast processing, a trained model, and more computational power than a simple machine-learning setup. Another limitation was that the music used for training came from multiple online sources, which made the data uneven in terms of sound quality and recording style. The paper did not test any small, controlled audio samples, so it was difficult to understand how individual features contribute to emotion.[9]

10. Research Gap

Across the studies reviewed, a common issue appears again and again. Most authors worked with large collections of professionally produced music, which often come from many different sources and styles. Because these recordings are not controlled, it becomes difficult to understand how individual features such as tempo, loudness, or spectral brightness influence the emotional impression of a track. Several papers also relied on deep-learning models that gave good accuracy but did not clearly explain why the prediction was made, making the systems hard to interpret. Very few studies created their own musical samples or tested emotions using simple, self-composed clips. This leaves a clear gap for experiments that use controlled music created specifically for analysis. In this project, short emotion-based clips are composed in LMMS so that tempo, pitch, and overall sound can be adjusted intentionally, making it easier to study how basic DSP features relate to different emotional states.

11. References:

[1] J. Kang and D. Herremans, “Are We There Yet? A Brief Survey of Music Emotion Prediction Datasets, Models and Outstanding Challenges,” *IEEE Trans. Affect. Comput.*, pp. 1–16, 2025, doi: 10.1109/TAFFC.2025.3583505.

[2] P. L. Louro, H. Redinho, R. Malheiro, R. P. Paiva, and R. Panda, “A Comparison Study of Deep Learning Methodologies for Music Emotion Recognition,” *Sensors*, vol. 24, no. 7, pp. 1–17, 2024, doi: 10.3390/s24072201.

[3] X. Jia, “Music Emotion Classification Method Based on Deep Learning and Explicit Sparse Attention Network,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3920663.

[4] J. Kang and D. Herremans, “Towards Unified Music Emotion Recognition across Dimensional and Categorical Models,” 2025, [Online]. Available: http://arxiv.org/abs/2502.03979

[5] R. Liyanarachchi, A. Joshi, and E. Meijering, “A Survey on Multimodal Music Emotion Recognition,” *ACM Comput. Surv.*, vol. ?, no. ?, 2025, [Online]. Available: http://arxiv.org/abs/2504.18799

[6] C. Huang and Q. Zhang, “Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect,” *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/3807666.

[7] Z. Huang, S. Ji, Z. Hu, C. Cai, J. Luo, and X. Yang, “ADFF: Attention Based Deep Feature Fusion Approach for Music Emotion Recognition,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2022–Septe, no. 1, pp. 4152–4156, 2022, doi: 10.21437/Interspeech.2022-726.

[8] J. S. Gomez-Canon *et al.*, “Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications,” *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 106–114, 2021, doi: 10.1109/MSP.2021.3106232.

[9] X. Hao, H. Li, and Y. Wen, “Real-time music emotion recognition based on multimodal fusion,” *Alexandria Eng. J.*, vol. 116, no. January, pp. 586–600, 2025, doi: 10.1016/j.aej.2024.12.060.