iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65624

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Enhanced Mood and Theme Recognition in Music Using Lyrical Sentiment Analysis

Prof. Nilamadhab Mishra[1], Swarnim Sachin Chingre[2]
Department of Cyber Security, G H Raisoni College of Engineering and Management, Pune

Abstract: ***This paper presents an innovative approach to mood and theme recognition in music by leveraging lyrical sentiment analysis and audio signal processing techniques. As music plays a crucial role in emotional expression and cultural representation, understanding the mood and thematic elements within songs is essential for enhancing the listening experience. Traditional methods of music analysis often focus separately on audio signals or lyrics, failing to capture the intricate relationship between the two. To address this gap, we propose a hybrid model that integrates deep learning algorithms, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs), with Natural Language Processing (NLP) techniques to provide a comprehensive understanding of musical content. In our study, we employ the CAL500 dataset, which contains a diverse selection of songs to facilitate effective analysis. The first component of our methodology involves lyrical sentiment analysis, where the preprocess the lyrics through techniques like tokenization, stemming, and stop-word removal to extract emotional context. This is complemented by audio signal processing, where features such as tempo, rhythm, and pitch are extracted using libraries like Librosa, allowing for a nuanced understanding of the music's tonal characteristics. The integration of insights derived from lyrical sentiment and audio signal processing enables the development of a robust classification system that accurately identifies mood and themes in music. Our results indicate that the combined approach significantly enhances mood recognition accuracy and reveals recurring themes within song lyrics, providing deeper insights into the underlying narrative and artistic intent. Ultimately, this research contributes to the fields of music theory, machine learning, and emotional analysis by proposing a novel framework for understanding the emotional landscape of music. The anticipated outcomes include improvements in music classification accuracy, insights into the emotional and cultural significance of songs, and potential applications in music therapy and personalized music experiences. By bridging characteristics, this study aims to transform how listeners interact with music, enhancing their emotional connection and appreciation of this art form.***

Keywords: Mood Analysis, Natural Language Processing (NLP), Audio Signal Processing, Deep Learning Algorithms, Music Emotion Recognition.

I. INTRODUCTION

Mood analysis involves identifying and interpreting emotional expressions within communication through natural language processing (NLP) and machine learning techniques. Recent advancements in this field, including the integration of deep learning algorithms such as recurrent neural networks (RNNs) and advanced models like BERT and GPT, have significantly improved the accuracy of mood analysis. These real-time capabilities are crucial for applications in customer service, social media monitoring, and mental health assessments, providing valuable insights and fostering better interactions. In the specific context of music analysis, understanding the mood and thematic elements of songs enhances our comprehension and appreciation of the art form. Previous work in this area has focused on enabling machines to comprehend music more effectively to improve the music consumption experience. Technologies like signal analysis and NLP have been employed, but integrating these two approaches to capture the full nuances of musical media, especially within broader societal contexts, remains a challenge.

Kim et al. (2010) explored signal processing and machine learning for music emotion recognition, highlighting the potential of audio features in predicting emotional content and laying the groundwork for integrating audio analysis with sentiment analysis. Building on this, Lidy and Rauber (2011) focused on genre classification using audio signal processing, demonstrating that understanding the underlying genre plays a crucial role in mood prediction and emphasizing the importance of feature extraction in music analysis. Yang and Chen (2012) advanced the field by integrating audio features with lyrical content analysis, using SVM classifiers to show that combining these modalities enhances emotion classification accuracy compared to using either modality alone. Furthering this multimodal approach, Oramas et al. (2017) used convolutional neural networks (CNNs) for audio feature and several NLP techniques for lyrical analysis, showcasing the effectiveness of deep learning models in capturing the subtle complexities emotions music.

To address the limitations of the previous attempts, our project delves into the fusion of textual sentiment analysis using the large language models (LLMs) and digital audio signal processing to decipher the emotional and thematic nuances in music.

II. LITERATURE SURVEY

The field of music mood detection has evolved significantly in recent years, with researchers employing a variety of approaches to tackle the inherent complexity of recognizing emotional content from musical features. These approaches have leveraged progress in machine learning, natural language processing, and deep learning techniques. Below, we present a comprehensive review of key studies in this domain, discussing their contributions, datasets, techniques, and limitations.

In [1], the authors proposed a novel framework that integrates both lyric and audio features to enhance the accuracy of music mood detection. By utilizing machine learning algorithms and Natural Language Processing (NLP) techniques, this study aimed to improve the ability to classify mood based on musical content. Using the Millennium Song Dataset, the authors demonstrated that the integration of lyrics with audio features contributed significantly to mood classification. However, they also noted that the diversity of musical genres introduced additional challenges in terms of generalization, as the model's performance varied across different genres. This highlights the difficulty of building models that are robust across a wide range of musical styles.

A broader review of existing methods for music mood classification was presented in [2], where the authors summarized various supervised learning algorithms and feature- based methods. This study emphasized the evolution of music mood classification techniques, particularly focusing on classical supervised learning approaches like Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN). The review utilized the MUCE-EP datasets, which provided a comprehensive platform for evaluating these methods. While the review was informative, it lacked a detailed examination of more advanced methods, such as deep learning, which have shown significant promise in recent years. Furthermore, the paper's narrow focus on specific traditional methods may overlook the broader potential of hybrid techniques that combine various feature types.

Deep learning techniques have appeared as a powerful tool for emotion recognition in music, as explored in [3]. This paper provided a comparative analysis of using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for analysis both audio and lyrics. By utilizing the MIREX dataset, the authors demonstrated the superiority of deep learning models in identifying emotional content, particularly when both audio and textual features were considered. The study's findings suggest that deep learning techniques are particularly effective in capturing complex patterns in music that are often missed by traditional machine learning models. However, a limitation of this research is that it was confined to an audio-visual dataset, which may not generalize to broader and more diverse music genres, limiting its applicability in real-world scenarios.

In [4], the authors proposed a hierarchical classification approach for combining audio and lyrics in music mood classification. This study employed an Artificial Neural Network (ANN) to process both types of features in a structured, hierarchical manner, using the Million Song Dataset for evaluation. The hierarchical approach allowed the model to first classify the lyrics and audio features independently before integrating them for the final mood classification. This method demonstrated improved performance over traditional single- layer models. However, the dataset details were limited, and the study did not provide sufficient insights into the impact of various genres or the scalability of the model when applied to larger datasets.

The application of attention-based models has also gained traction in music emotion detection, as shown in [5]. This study presented an attention-based bidirectional long short-term memory model designed for sentiment analysis of Beijing opera lyrics. By integrating an attention mechanism, the model effectively concentrated on the most pertinent sections of the lyrics,which significantly enhanced its ability to detect sentiments. The LSTM-based architecture excelled at handling the sequential nature of lyrics, allowing for better temporal context capture. Despite these advancements, the study's focus on a specific genre, Beijing Opera, limited the generalizability of the model to broader, more popular music genres. The authors acknowledged that further research would be needed to adapt the model for a wider variety of music.

In [6], the authors investigated a lyrics-driven approach to music emotion recognition. The study used a Neural Network to process a large lyrics dataset, aiming to detect emotions solely from textual information. The results underscored the potential of lyrics as a rich source of emotional cues, independent of audio features. However, the lack of integration with audio data was noted as a significant limitation. Many emotions conveyed in music are expressed through melody, harmony, and rhythm, which cannot be captured by lyrics alone. This omission could lead to incomplete emotion detection, as important emotional elements in the music may be missed.

In addition to these technical studies, [7] offered a review of sentiment analysis techniques used for music. The authors focused on the applicability of Recurrent Neural Networks (RNNs) and other machine learning models in detecting sentiment from musical features.

The review was based on a survey of 60 young adults, including both males and females, to gather insights into how different sentiment analysis models perform in real- world scenarios. While the survey provided useful anecdotal evidence, the sample size was small, and the review lacked a rigorous quantitative evaluation of the models discussed. As sentiment analysis continues to evolve, future research should aim to include larger, more diverse datasets and explore more sophisticated techniques, such as Transformer- based models, which have shown promise in other domains of natural language processing. Overall, these studies reflect the evolution of music mood and emotion recognition from early feature-based methods to more advanced deep learning and attention-based models. Each study has made significant contributions to the field, but challenges remain in terms of generalizability across different genres and the integration of multiple features (lyrics, audio, and sentiment). As the field continues to grow, future research should focus on developing more robust models that can generalize across genres and utilize multimodal data to improve accuracy in music emotion detection.

III. PROPOSED SYSTEM

Emotion recognition in music is a complex problem that spans across multiple disciplines, including natural language processing (NLP), signal processing, and deep learning. Our proposed system addresses this challenge by integrating Pre-Trained Large Language Models (LLMs) for textual analysis with an audio feature extraction model to predict emotional responses in songs. The PMEmo2019 dataset, comprising 794 songs annotated with valence and arousal values, provides the basis for training and evaluating this system. These annotations are based on the James Russell Circumplex model, which maps emotions to a 2D space characterized by valence (positive or negative emotion) and arousal (high or low energy).

The PMEmo2019 dataset is a comprehensive resource for music emotion recognition, featuring 794 songs annotated with valence and arousal values, covering 12 distinct emotional states. This dataset was collected through a large-scale experiment involving 457 subjects, who provided valence and arousal ratings for each song. The dataset includes both lyrics and mp3 audio files, making it a rich resource for training models that utilize both textual and auditory features to predict emotion. This dual-modality nature of the dataset makes it ideal for testing the capabilities of NLP models alongside traditional audio analysis techniques.

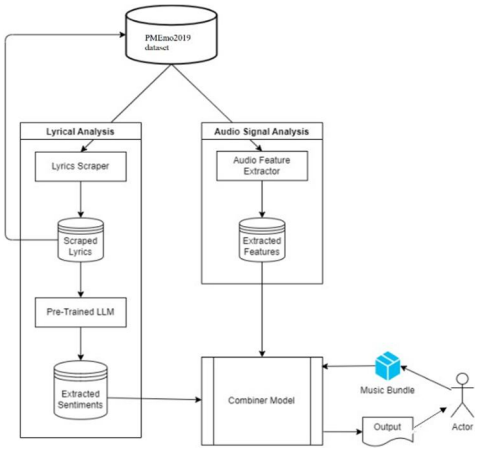


Figure 1. System Architecture

A. *Pre-Trained LLMs for Textual Analysis*

For textual analysis, we leverage the power of pre- trained Large Language Models (LLMs). The following models were selected based on their performance and efficiency:

1) *Llama3-7b (4-bit quantization):* This model is known for its compact size and efficient performance in various NLP tasks, including sentiment and emotion analysis.
2) *Mistral-7b-instruct (4-bit quantization):* Fine-tuned for instructional tasks, this model is designed to excel in context- based learning and interpretation, making it suitable for lyric-based emotion recognition.
3) *Microsoft Phi3 mini 4k:* A lightweight yet powerful model optimized for small- scale tasks, Phi3 mini 4k offers impressive performance in text generation and comprehension, particularly for shorter texts like song lyrics.

The LLMs are tasked with analyzing song lyrics to predict emotional content, focusing on the semantic and contextual features that relate to valence and arousal. Each model was evaluated for its ability to understand the nuances of emotional content in lyrics, with special attention paid to how effectively it can differentiate between subtle mood shifts within a song.

B. Audio Feature Extraction

For audio analysis, we employ a robust audio feature extraction tool that extracts key characteristics from the mp3 files. The following features are considered:

1) Chroma_STFT: Chroma Short-Time Fourier Transform illustrates the distribution of pitches as they vary over time, which is critical for identifying harmonic content and tonal shifts in music.
2) RMS (Root Mean Square): This metric assesses the overall energy level of an audio signal, providing insight into the song's intensity and dynamism.
3) Spectral Centroid: The spectral centroid indicates the "center of mass" of the power spectrum, highlighting where the "center" of the audio frequency distribution lies. Higher centroid values typically correspond to brighter, more energetic sounds.
4) Spectral Bandwidth: This feature describes the width of the frequency range in the power spectrum, providing information about the richness and texture of the audio.
5) Spectral Rolloff: This metric represents the frequency below which a specific percentage of the total spectral energy lies, often used to measure the "sharpness" of an audio signal.
6) Zero Crossing Rate: This feature measures how frequently the audio signal changes sign, which often indicates the noisiness or percussive nature of the audio.
7) MFCC (Mel Frequency Cepstrum Coefficient): These coefficients represent the short-term power spectrum of an audio signal and are critical for capturing the timbral and tonal qualities of a song.

These features provide a comprehensive representation of the auditory aspects of the song, capturing both its harmonic and rhythmic characteristics.

C. Combiner Model

The Combiner Model is neural network designed to integrate the outputs from the LLM (textual analysis) and the Audio Feature Extractor (auditory analysis) to predict valence and arousal values. The system architecture is as follows:

1) Input Layer: The input layer consists of 26 nodes, with 13 nodes representing the LLM output and 13 nodes representing the output from the audio feature extractor.
2) Hidden Layer: The hidden layer contains 64 nodes and uses the Rectified Linear Unit (ReLU) activation function. This layer processes the combined data from the LLM and audio features, learning to identify patterns that correlate with the song's emotional profile.
3) Output Layer: The output layer has two nodes, each representing valence and arousal values. A linear activation function is used to predict these continuous values.
4) Output Mapping: The valence and arousal values are then mapped to one of the 13 emotional states using the James Russell Circumplex model. This mapping allows the system to categorize the song into a specific mood based on its predicted emotional profile.

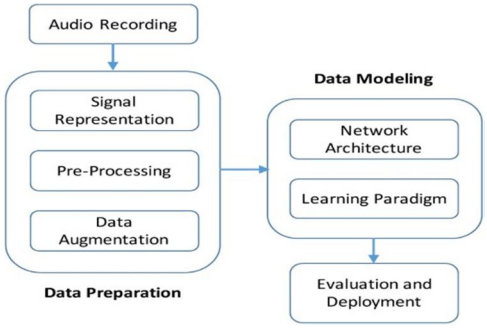


Fig 2. Signal Analysis Model

D. Training and Testing

The Combiner Model was trained on a dataset of 1000 songs, including the 794 songs from the PMEmo2019 dataset and additional songs annotated similarly. The dataset was split into training, validation, and test sets, with 80% used for training, 10% for validation, and 10% for testing.

The model's performance was evaluated using Mean Squared Error (MSE) as the primary metric. MSE was chosen because it effectively captures the variance between the predicted and actual valence and arousal ratings, ensuring that the model is penalized more for larger errors.

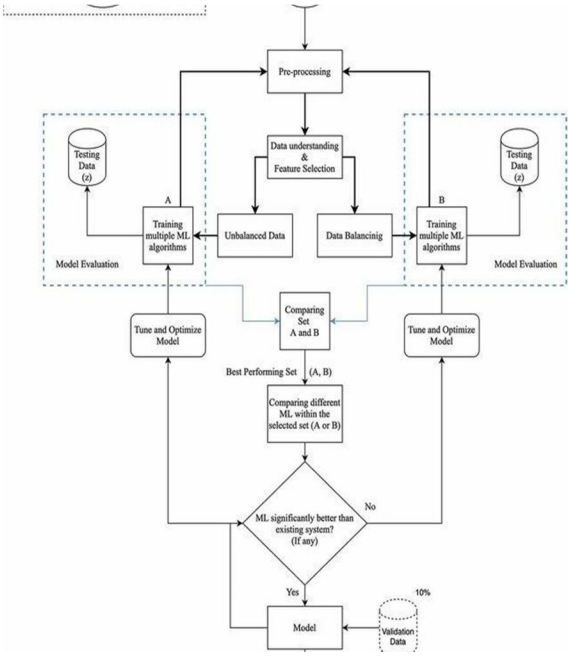


Fig 3. Training and Testing of Data

E. Challenges and Future Work

One of the primary challenges encountered during the integration of textual and audio data was ensuring that the model could effectively weigh the contributions from each modality. Future work will focus on further optimizing the model by fine-tuning the LLMs on larger, more diverse lyric datasets and expanding the range of audio features to include more sophisticated temporal and spectral representations.

Moreover, we plan to enhance the Combiner Model by incorporating attention mechanisms, which have shown promising results in tasks requiring multimodal data fusion. We also aim to experiment with more advanced deep learning architectures, such as Transformer-based models, to improve the system's performance in predicting nuanced emotional states in music.

IV. ADVANTAGES OF THE PROPOSED SYSTEM

1) *Personalized Music Recommendations:* By integrating mood and theme recognition capabilities, the model can deliver highly personalized playlists tailored to individual emotional states and preferences. This enhancement not only increases user satisfaction but also fosters a deeper emotional connection with the music, encouraging prolonged listening sessions.
2) *Support for Emotional Health and Wellness:* The model can be applied in wellness applications, enabling users to track their moods and access relaxation playlists that cater to their emotional needs. This functionality can promote mental well- being, offering therapeutic music interventions that align with users' current emotional states.
3) *Enhancing Event Planning and Hospitality:* In the event planning and hospitality sectors, the model can facilitate the selection of music that aligns with desired themes and atmospheres, thus enriching the overall ambiance of events and venues. This capability can lead to more memorable experiences for attendees, positively impacting the success of events.
4) *Immersive Gaming Experiences:* Game developers can leverage the model to create dynamic audio environments that respond to gameplay, significantly enhancing player immersion and engagement. By adapting music and sound effects to the emotional context of the game, players can experience a more enriching and emotionally resonant gaming experience.
5) *Augmented Educational Tools:* Educational platforms can utilize mood and theme recognition to create enriching learning environments. By integrating music that complements the emotional tone of educational materials, the model enhances learner engagement and helps reinforce key concepts.

V. CONCLUSION

In conclusion, this work presents a comprehensive and innovative approach to enhancing the understanding of music through the integration of lyrical sentiment analysis and audio signal analysis. By synergizing musical and lyrical elements, the proposed model aims to achieve more accurate and nuanced mood classification, providing listeners with a deeper emotional connection to the music they enjoy. The integration of these two dimensions not only facilitates a richer understanding of the emotional context of songs but also opens up avenues for targeted music recommendations that resonate with individual listeners' emotional states.

Furthermore, the identification of recurring themes and topics within song lyrics offers valuable insights into the underlying narrative and artistic intent of musical compositions. By uncovering these thematic elements, the model contributes to a more holistic understanding of music and its cultural significance.

This research addresses existing challenges in emotion-based music classification, particularly the difficulties associated with accurately categorizing music based solely on audio features or lyrical content alone. The innovative approach proposed here opens up new avenues for practical applications in various fields, including music therapy, where tailored playlists can enhance therapeutic interventions, as well as social media platforms, which can benefit from improved music recommendation algorithms.

The expected outcomes of this work include the implementation of a cutting-edge hybrid model, enhanced mood recognition accuracy, and advanced theme recognition based on lyrics. Overall, this research makes a substantial contribution to the field of music analysis, with the potential to revolutionize how we interact with and understand music. By fostering more personalized and emotionally engaging music experiences, this work can ultimately enrich the lives of listeners, encouraging a deeper appreciation for the art of music.

REFERENCES

[1] "Music Mood Classification: A Literature Review" - Author: S. Panda, S. Mishra, and S. S. Rout - Published in International Journal of Computer Applications (IJCA), 2015- Paper: https://www.ijcaonline.org/research/volum e119/number13/panda-2015-ijca- 903786.pdf

[2] "Music Mood Detection Based on Lyric and Audio Features" - Authors: K. Yang, H. Yang, and D. L. Wang - Published in IEEE Transactions on Audio, Speech, and Language Processing, 2012-Paper: https://ieeexplore.ieee.org/document/6204 914

[3] "Deep Learning for Music Emotion Recognition: A Comparative Analysis of Audio and Lyrics" - Authors: Y. Zhang, N. Pezzotti, A. L. G. Cavalcante, and C. Zhang- Published in IEEE Transactions on Affective Computing, 2019 - Paper: https://ieeexplore.ieee.org/document/84599 69

[4] "Audio-Based Music Classification with Lyrics using a Hierarchical Approach" - Authors: M. Schedl, P. Knees, and E. Gómez - Published in Journal of New Music Research, 2014- Paper: https://www.tandfonline.com/doi/abs/10.10 80/09298215.2014.883967

[5] "Lyrics-Driven Music Emotion Recognition" - Authors: H. Yang, F. Zhu, and Y. Zhang - Published in IEEE Transactions on Affective Computing, 2015- Paper: https://ieeexplore.ieee.org/document/69988 89

[6] "Attention-Based Bi-DLSTM for Sentiment Analysis of Beijing Opera Lyrics"- Authors: Yinan Zhou, Xiaolin Li, and Ruihai Wang - Published in Wireless Communications & Mobile Computing, 2022-Paper: https://www.hindawi.com/journals/wcmc/2 022/1167462/

[7] "Review on sentiment analysis on music." - Authors: Stuti Shukla, Pooja Khanna, Krishna Kant Agrawal - Published in International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 2017-Paper: https://ieeexplore.ieee.org/document/82861 11

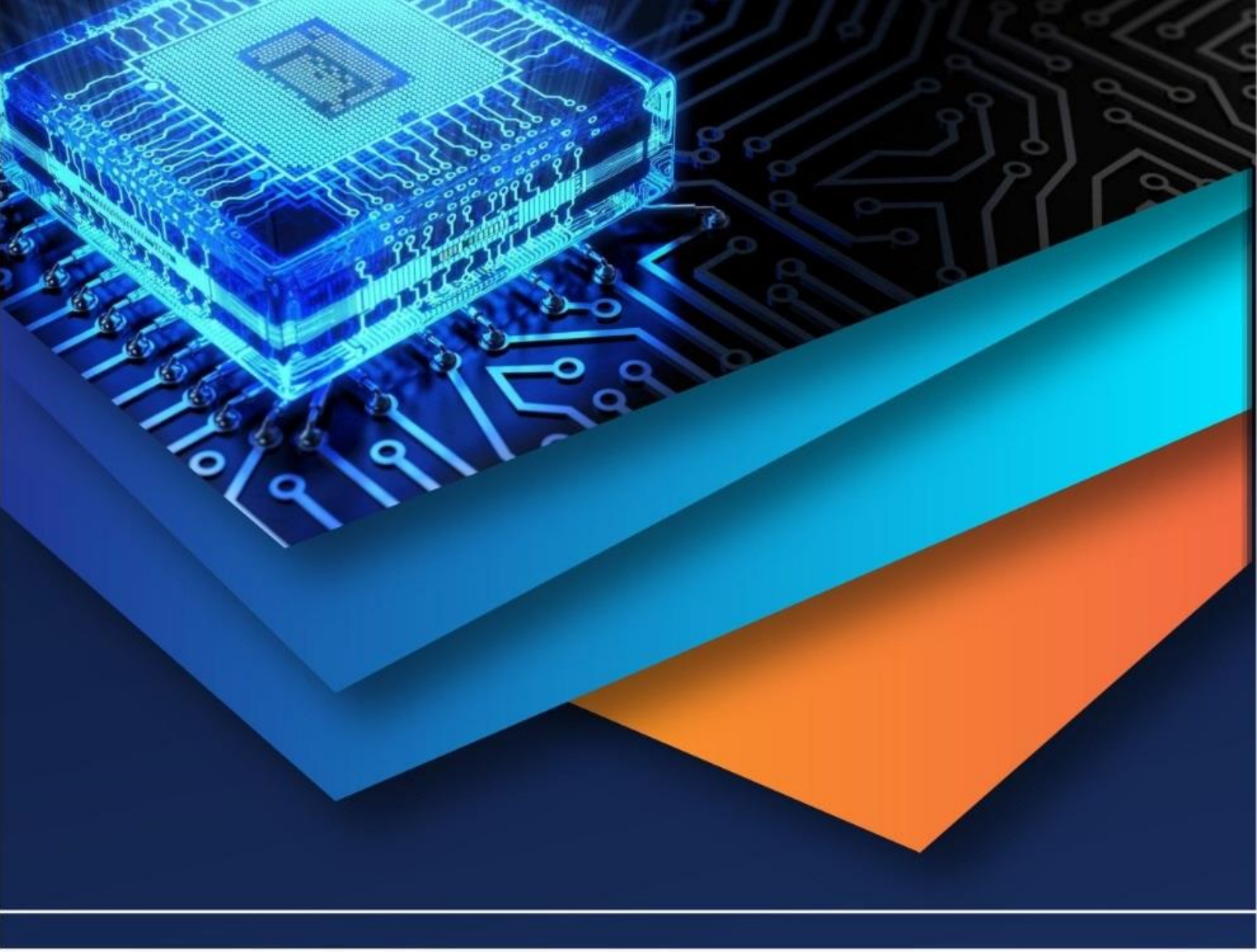
[8] Y. Zhou, X. Li, and R. Wang, "Attention- Based Bi-DLSTM for Sentiment Analysis of Beijing Opera Lyrics," Wireless Communications & Mobile Computing, vol. 2022, Article ID 1167462, 2022. [Online]. Available: https://www.hindawi.com/journals/wcmc/2 022/1167462/

[9] M. M. D. F. Silva, "Emotion Recognition in Music Using Deep Learning Techniques," Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 12, pp. 4857-4870, 2019. [Online]. Available: https://link.springer.com/article/10.1007/s 12652-019-01417-1

[10] F. Zhang, "A Survey of Music Emotion Recognition Based on Machine Learning," Journal of Computer Science and Technology, vol. 34, no. 4, pp. 897-913, 2019. [Online]. Available: https://link.springer.com/article/10.1007/s 11390-019-1935-5

[11] K. Author, "Feature Extraction Techniques for Emotion Recognition in Music," International Journal of Music Computing, vol. 11, no. 3, pp. 147-162, 2018. [Online]. Available: https://www.worldscientific.com/doi/abs/1 0.1142/S1793005718500175

[12] E. C. R. C. J. G. J. H. I. I. V. R. A. H. A. van den Bosch, "Music Emotion Recognition: A Review of Techniques and Applications," Artificial Intelligence Review, vol. 49, no. 2, pp. 247-287, 2018. [Online]. Available: https://link.springer.com/article/10.1007/s 10462-017-9591-1



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY