International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 3
ISSN: 2088-5334

Emotion Recognition and Multi-class Classification in Music with MFCC and Machine Learning

Gilsang Yoo [a,*], Sungdae Hong [b], Hyeocheol Kim [a]

[a] *Creative Informatics and Computing Institute, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea*
[b] *Division of Design, Seokyeong University, 124 Seogyeong-ro Seongbuk-gu Seoul, 02173, Republic of Korea*
*Corresponding author: *ksyoo@korea.ac.kr*

Abstract—**Background music in OTT services significantly enhances narratives and conveys emotions, yet users with hearing impairments might not fully experience this emotional context. This paper illuminates the pivotal role of background music in user engagement on OTT platforms. It introduces a novel system designed to mitigate the challenges the hearing-impaired face in appreciating the emotional nuances of music. This system adeptly identifies the mood of background music and translates it into textual subtitles, making emotional content accessible to all users. The proposed method extracts key audio features, including Mel Frequency Cepstral Coefficients (MFCC), Root Mean Square (RMS), and MEL Spectrograms. It then harnesses the power of leading machine learning algorithms Logistic Regression, Random Forest, AdaBoost, and Support Vector Classification (SVC) to analyze the emotional traits embedded in the music and accurately identify its sentiment. Among these, the Random Forest algorithm, applied to MFCC features, demonstrated exceptional accuracy, reaching 94.8% in our tests. The significance of this technology extends beyond mere feature identification; it promises to revolutionize the accessibility of multimedia content. By automatically generating emotionally resonant subtitles, this system can enrich the viewing experience for all, particularly those with hearing impairments. This advancement not only underscores the critical role of music in storytelling and emotional engagement but also highlights the vast potential of machine learning in enhancing the inclusivity and enjoyment of digital entertainment across diverse audiences.**

Keywords—**Emotion recognition; multi-class classification; machine learning; Mel spectrograms.**

Manuscript received 11 Sep. 2023; revised 29 Nov. 2023; accepted 7 Jan. 2024. Date of publication 30 Jun. 2024.

I. Introduction

In contemporary society, multimedia content, mainly OTT services, has become a vital component of daily life, offering diverse video content that enriches user experiences [1]. These services enhance narratives and convey emotions through background music [2]. However, users with auditory limitations, such as hearing impairment, may find it challenging to fully grasp the emotional context conveyed by these musical elements [3]–[5]. To address this issue, this study proposes a system within OTT services that automatically recognizes the mood of background music in video content and converts it into textual subtitles. This system aims to improve the accessibility and inclusiveness of multimedia content, enabling all users, including those with hearing impairments, to have a more enriching multimedia experience.

The objectives of this research are to develop techniques for accurately extracting and analyzing the characteristics of background music in video content, explore methods for effectively converting the mood and emotional qualities of music into text, and evaluate whether the generated subtitles can effectively convey the emotional context of the video to hearing-impaired users. To achieve these goals, the study employs machine learning to analyze background music's mood and emotional qualities and convert them into text. Fundamental techniques used in this process include Mel-Frequency Cepstral Coefficients, Root Mean Square, and MEL Spectrogram analysis. These techniques extract meaningful features from audio signals and then classify and textualize the music's mood through various machine learning algorithms, including Random Forest, Regression Analysis, AdaBoost, and SVC (Support Vector Classification). The application of machine learning for background music analysis represents a significant advancement in this field, contributing significantly to the understanding and processing of multimedia content.

This research is expected to enable all users, including those with hearing impairments, to experience multimedia content fully. By enhancing content accessibility and inclusiveness, this research could establish new standards for producing and delivering video content. The next part of this paper, section 2, reviews existing background music recognition and mood analysis research. It presents the technical details and algorithms used in this study and the system's implementation method. The experimental results assess the system's performance and potential benefits for hearing-impaired users. Finally, section 4 reveals the implications of the research findings and future research directions.

II. Materials and Methods

A. MFCC Technique for Background Music Processing

The Mel-Frequency Cepstral Coefficients technique plays a pivotal role in speech recognition and music information retrieval, showcasing notable success in processing background music. Developed based on the human auditory system's use of the Mel scale rather than a linear frequency scale to perceive sounds, MFCC effectively extracts essential features such as timbre, pitch, and rhythm from musical signals, mirroring human auditory characteristics [6].

Initially utilized primarily in feature extraction for speech recognition systems, MFCC has expanded its application to analyzing background music characteristics through the work of various researchers, finding widespread use in areas such as music genre classification, emotion analysis, and music recommendation systems [7].

Furthermore, MFCC-based background music processing techniques have been applied to analyze the mood and emotions of background music in films and videos [8]–[9], [10]. By classifying the emotional states of movie background music based on features extracted through MFCC, it is possible to analyze the overall emotional flow of the movie. These studies highlight MFCC's potential as a tool for music information retrieval and multimedia content analysis and understanding [11]–[13]. Building on these previous studies, this research aims to utilize MFCC technology for processing background music in video content, automatically developing a system to textualize music's mood and emotional characteristics. This study aims to extend existing MFCC-based music processing techniques to enable a broader audience, including users with hearing impairments, to understand the musical elements of video content.

B. Mel Spectrogram Analysis

The Mel Spectrogram is an essential tool in audio signal analysis, widely utilized particularly in music and speech processing. This technique visualizes the changes in frequency over time, aiding in understanding the complex structures within audio signals. Designed to mimic the characteristics of human hearing, the Mel scale processes frequencies akin to how humans perceive sound, sharing similarities with MFCC analysis but emphasizing the visual representation of audio signals [14]–[15].

Recent studies have demonstrated the effectiveness of Mel Spectrograms when used in conjunction with deep learning models, especially Convolutional Neural Networks (CNN), for automatic classification of music genres, music recommendation systems, and emotion analysis based on music. For instance, CNN models utilizing Mel Spectrograms as input have been developed to classify music genres with high accuracy [16]–[17]. This approach allows deep learning models to effectively learn the complex features of audio signals, yielding superior results compared to traditional manual feature extraction methods.

Moreover, due to their ability to capture acceptable audio signal variations, Mel Spectrograms are also valuable for analyzing the subtle emotional nuances within music [18]–[19]. This facilitates a more sophisticated understanding of the emotional impact of background music in films or videos on viewers. Thus, Mel Spectrogram analysis is a powerful tool for effectively visualizing and analyzing the complex features of audio signals, offering various applications in music processing and analysis. This study aims to leverage the advantages of Mel Spectrograms to analyze the mood of background music in video content and convert it into subtitle information for users with hearing impairments, presenting a novel approach to this challenge.

C. Machine Learning for Background Music Analysis

Background music analysis has established itself as one of the crucial elements in understanding and processing multimedia content, with the application of machine learning emerging as a significant research focus in this area. Leveraging machine learning techniques allows researchers to automatically identify and classify the complex characteristics of background music, enhancing the content's emotional ambiance and viewer experience [20].

In early studies, machine learning was primarily applied to simple tasks, such as classifying music genres. Based on these attributes, these studies quantified various musical attributes (e.g., rhythm, melody, harmony) and trained classification algorithms (e.g., decision trees, K-nearest neighbors). While this approach proved helpful in identifying the musical features of specific genres, it faced limitations when analyzing more complex characteristics, such as the subtle emotional nuances of background music.

Recent studies have started applying more advanced machine learning models, particularly techniques like deep learning, to background music analysis to overcome these limitations. For instance, CNNs and Recurrent Neural Networks (RNNs) are particularly effective in learning temporal changes and patterns in audio signals, enabling the practical analysis of complex characteristics like the emotional mood of background music [21]–[22]. Furthermore, machine learning techniques have been utilized to analyze the interactions between background music and video content [23]–[29]. This approach has enabled a more accurate understanding of the impact of background music on viewers, allowing content creators to design more effective emotional experiences.

D. Data Exploration

The dataset is composed as follows: The dataset, which is publicly available on Kaggle (https://www.kaggle.com/datasets/dikshashri13702/features-music-mood-classification/data), consists of 2500.wav files that are labeled with five emotions:

- Aggressive (500 files)
- Dramatic (500 files)
- Happy (500 files)
- Romantic (500 files)
- Sad (500 files)

The emotional music is composed of 5-second segments, and the signal waveform for each piece of music data is shown in Fig. 1.

E. Feature Extraction

Audio features such as Mel Frequency Cepstral Coefficients, Root Mean Square, and Mel Spectrograms are extracted in the feature extraction stage. These features numerically represent various acoustic properties of music and are used as inputs for machine learning models. The computational methods for each feature at this stage are as follows:

1) Fourier Transform (FFT): Fourier Transform is performed for each frame to convert it into the frequency domain. The FFT is conducted to transform the time-domain signal into the frequency domain and extract the energy distribution in the spectrum. The amplitude spectrum of the signal is then converted into the energy spectrum using Equation (1). The window size used for calculating the FFT (Fast Fourier Transform) is set to 4096. A larger value increases the frequency resolution but decreases the time resolution.

$$X(k) = |FFT(x[n])|^2 \qquad (1)$$

The equation $X[k] = |FFT(x[n])|^2$ represents the power spectrum of a discrete signal. Here, *X[k]* denotes the energy at the *k* th frequency bin, while *FFT(x[n])* stands for the Fourier Transform of the signal *x[n]*. The modulus squared *|FFT(x[n])|^2* is used to calculate the power of the given frequency component in the signal. This is commonly used in signal processing to analyze the frequency content of signals.

2) Application of Mel Filter Bank: The human ear is more sensitive to lower frequencies than higher frequencies, a characteristic that the Mel filter bank considers. As frequency increases, the bandwidth of Mel-filters broadens to extract sufficient energy information from the lower frequency bands. The boundaries of each filter are calculated using a fixed equation between frequency and Mel frequency. Post-processing tasks, including logarithmic multiplication and Discrete Cosine Transformation (DCT), divide the frequency domain into several bands according to the Mel scale and calculate the energy in each band, transforming the filtered signals into MFCC features.

3) Log Energy: The logarithm of the energy is taken for each band.

4) Discrete Cosine Transform (DCT): The log energy spectrum is subjected to the DCT to calculate the MFCCs as illustrated in Equation (2).

$$\mathrm{MFCC[i]} = \sum_{\mathrm{n=0}}^{\mathrm{N-1}} \log(\mathrm{E_n})\cos\left(\left(\frac{\mathrm{i(n-0.5)\pi}}{\mathrm{N}}\right)\right) \qquad (2)$$

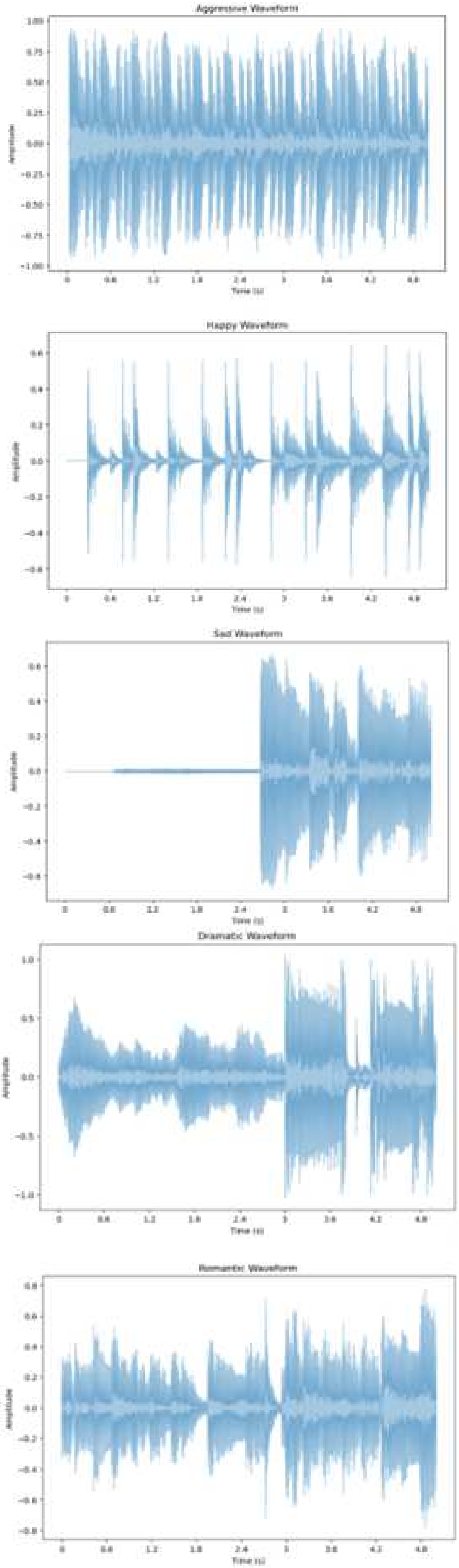

Fig. 1 The waveform of the signal for each music data

The equation provided is the formula for calculating the Mel Frequency Cepstral Coefficients (MFCCs). Here, *MFCC[i]* represents the *i*th cepstral coefficient, E_n denotes the log energy of the nth Mel filter bank channel, and N is the total number of Mel filter bank channels. The expression $log(E_n)$ indicates the logarithm of the energy, which is used to capture the non-linear human ear perception of sound. The cosine term is part of the Discrete Cosine Transform (DCT), which is applied to the log Mel spectrum, and *i* is the index of the MFCC. This transformation from the log Mel spectrum to the cepstral domain helps to decorrelate the signal and compresses the spectrum, resulting in a representation that can be effectively used in various audio processing tasks, particularly for voice and speech analysis. In the final analysis, *MFCC[i]* is stored in a 2D array containing 40 MFCC features calculated for each frame of the audio signal. Each column of the array represents a single frame of the audio signal, while each row corresponds to one of the MFCC coefficients for that frame. The transformed MFCC results effectively capture the timbre of the audio signal, as illustrated in Fig. 2., and are utilized in applications such as speech recognition and music classification.

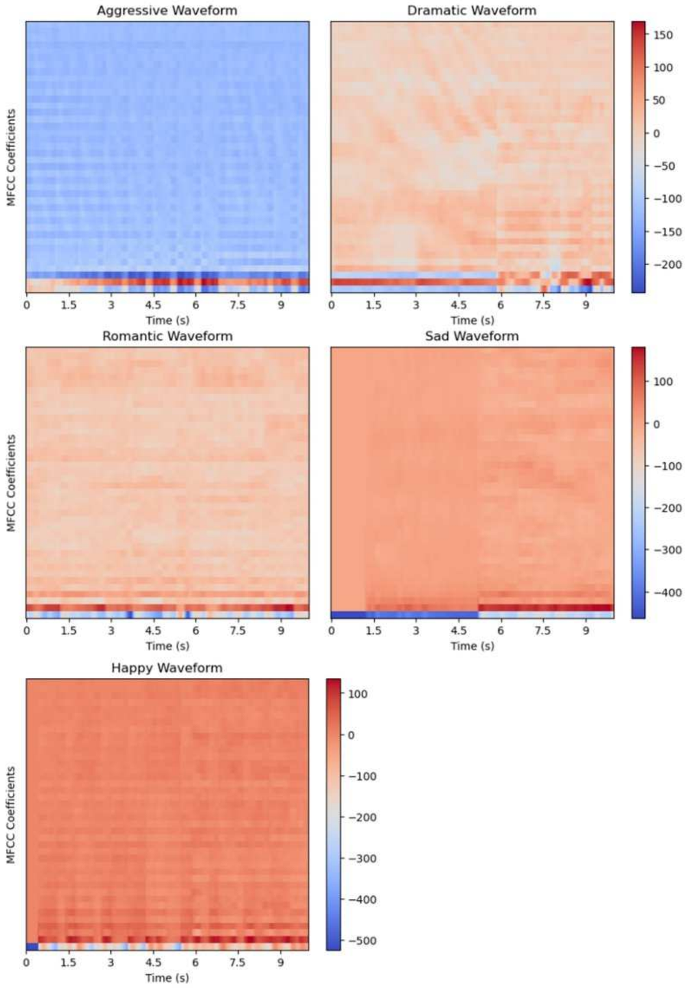

Fig. 2 The transformed MFCC results

5) RMS analysis: RMS, Delta RMSE, and Energy Novelty analyses are standard methods used in audio signal processing, especially in music and sound analysis. Each method examines specific aspects of a signal, helping to understand its characteristics:

- Root Mean Square: RMS is used to represent the average power of a signal. In audio, the RMS value roughly indicates the "loudness" or volume level of the signal. It's calculated by squaring all the sample values of the signal, averaging them, and then taking the square root of that average. RMS analysis is useful for understanding the overall energy level of audio and comparing the loudness between different audio clips or tracks.
- Delta RMSE (Root Mean Square Error): Delta RMSE represents the rate of change of RMS values over time. It can detect changes in energy levels in an audio signal and is particularly useful for analyzing dynamic range variations in music or audio clips. For instance, strong beats or sudden increases in loudness in music can manifest as sharp changes in Delta RMSE values.
- Energy Novelty: Energy Novelty analysis is used to find new points of energy change, or "novelty points," within an audio signal. It's primarily used in music structure analysis to identify significant changes within a track. Energy Novelty is obtained by calculating the energy of the signal over short periods and analyzing changes in this energy level. If the rate of change exceeds a certain threshold, that point can be considered a new change in energy.

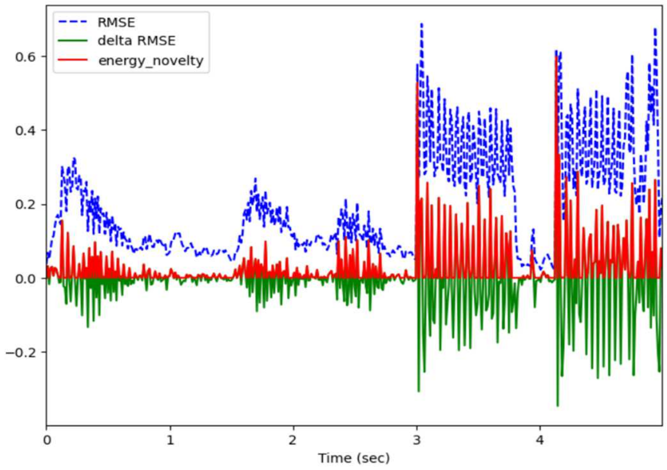

Fig. 3 Results of RMS Analysis for Dramatic Music

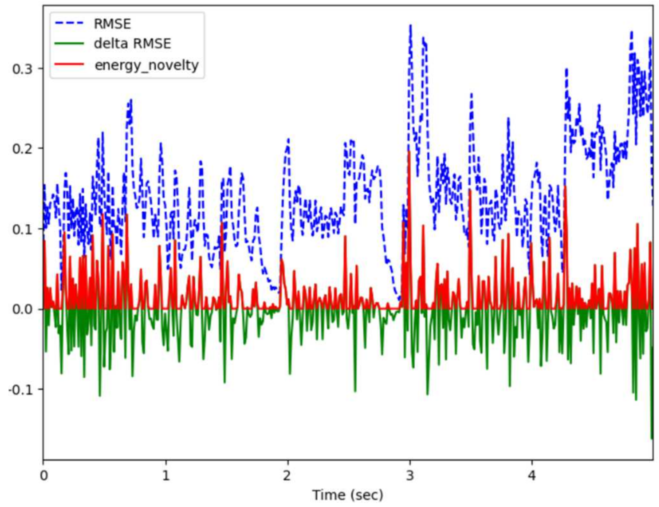

Fig. 4 Results of RMS Analysis for Romantic Music

Fig. 3. presents the results of the RMS, delta RMSE, and Energy novelty analyses for Dramatic music, while Fig. 4. shows the analysis results for Romantic music.

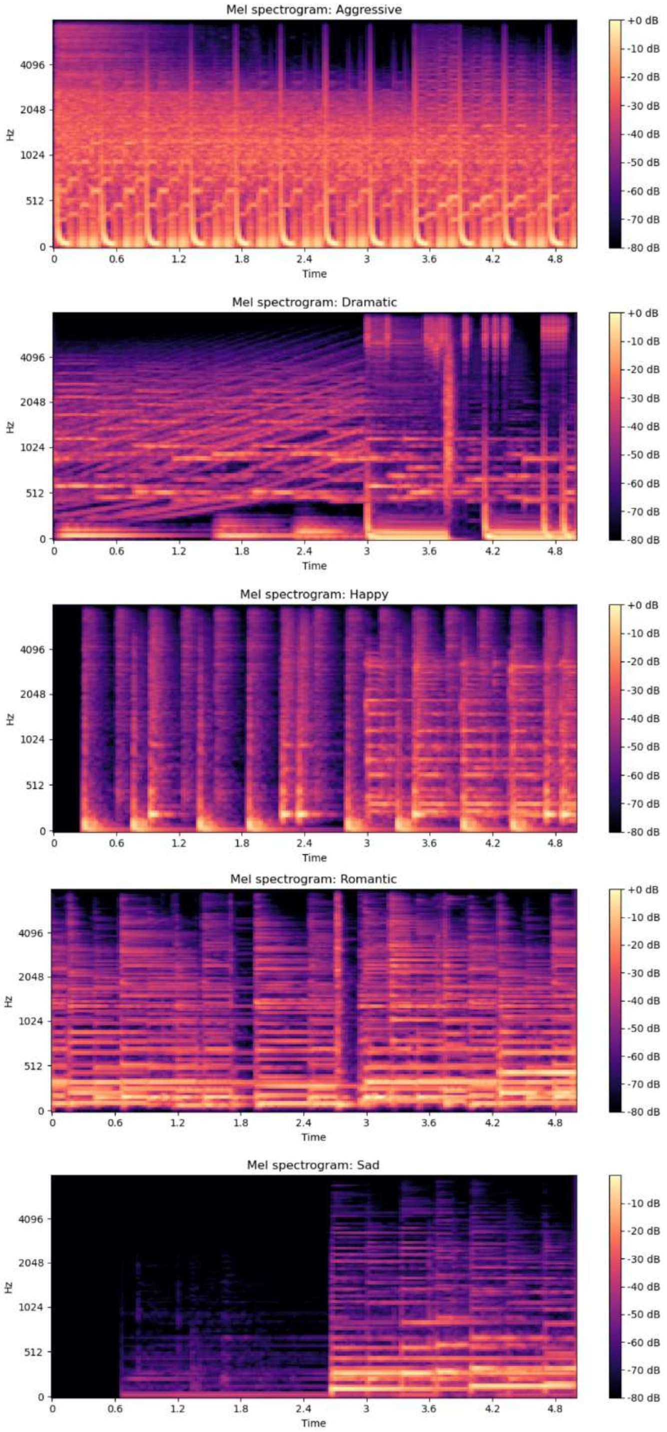

Fig. 5 Representative Mel Spectrogram Results for Each Piece of Music

F. *Quantification of Acoustic Properties*

The Mel Spectrogram is pivotal in machine learning, particularly in audio-related deep learning applications. It effectively extracts significant audio features through a Mel scale that mimics the human auditory system, offering these features in a format suitable for machine learning models to learn from. Transforming complex audio signals into a two-dimensional image format simplifies the high-dimensional complexity of data into a visually and computationally manageable form. This transformation also ensures high compatibility with powerful deep learning models such as Convolutional Neural Networks for image processing. Moreover, the Mel Spectrogram captures the temporal dynamics of audio signals, effectively conveying critical information about frequency variations over time to machine learning models. Due to these characteristics, the Mel Spectrogram is widely utilized as a crucial tool in machine learning and deep learning research related to audio analysis. The Mel Spectrogram can be obtained by arranging the energy of the Mel filter banks acquired during the MFCC calculation process along the time axis. Representative Mel Spectrogram results for each piece of music are depicted in Fig. 5.

III. Results and Discussion

Based on the extracted features, simulations were conducted using Logistic Regression, Random Forest, Support Vector Classification (SVC), and AdaBoost models. The models were trained using the training data and evaluated using the validation data. Where necessary, adjustments to the model's structure or optimization of hyperparameters were performed to enhance performance. The experimental results for each model are as follows.

A. *Logistic Regression Model*

Table 1 [30]–[33] displays the logistic regression model's performance metrics for a multi-class classification task. These metrics indicate the model's efficacy in predicting each class based on precision, recall, and f1-score [34].

- Aggressive: This class has the highest precision and recall at 96.5%, suggesting that the model predicts this class with excellent accuracy and reliability.
- Dramatic: Precision and recall are slightly lower than Class 0 at 91.6% and 92.2%, respectively, but still indicate high performance.
- Happy: With a precision of 93.9% and a recall of 96.0%, the model can identify this class correctly.
- Romantic: Presents the lowest precision and recall, at 83.6% and 83.0%, respectively, highlighting it as the most challenging class for the model to predict accurately.
- Sad: Scores are decent, with a precision of 85.7% and a recall of 83.8%. While not as high as other classes, the performance is satisfactory.

The model's overall accuracy is 90.3%, with the macro average and weighted average of precision, recall, and the f1-score closely matching this figure. This reflects a well-balanced performance across the classes, though there is room for improvement, particularly for Romantic and Sad, which show lower performance.

TABLE I
EVALUATION RESULTS OF THE LOGISTIC REGRESSION MODEL

Class	Precision	Recall	F1-Score
Aggressive	0.965	0.965	0.965
Dramatic	0.916	0.922	0.919
Happy	0.939	0.96	0.949
Romantic	0.836	0.83	0.833
Sad	0.857	0.838	0.847
accuracy	-	-	0.903
macro avg	0.902	0.903	0.903
weighted avg	0.902	0.903	0.903

B. *Random Forest Model*

The Random Forest model is an ensemble classification and regression technique that utilizes a collection of decision trees. Each tree in a Random Forest is trained on a random subset of the data, and the final prediction is made by aggregating (typically by voting for classification) the predictions of individual trees [35]–[40]. Table 2 shows the classification results for a Random Forest model, covering five different labels: Aggressive, Dramatic, Happy, Romantic, and Sad. Here's a breakdown of the performance metrics:

- Aggressive: It shows exceptional precision at 97.1% and perfect recall, meaning every instance of Aggressive in the test set was correctly identified. The F1 Score is 98.5%, indicating an excellent balance between precision and recall.
- Dramatic: High precision and recall, 96% and 95%, respectively, leading to a very high F1-Score of 95.5%, reflecting strong classification performance for this label.
- Happy: Achieved perfect precision, indicating that every prediction made as Happy was correct. The recall of 96% suggests that most, but not all, Happy instances were captured. The F1-Score is 98%, which is outstanding.
- Romantic: This label has lower precision at 87.6% but a higher recall of 92%, suggesting some false positives in the predictions. The F1-Score is 89.8%, the weakest among the labels, but still suggests good performance.
- Sad: It presents strong metrics, with precision at 93.8% and recall at 91%, resulting in a robust F1-Score of 92.4%.

The model's accuracy across all labels is 94.8%, indicating that the model correctly predicts the label 94.8% of the time across the dataset. The results reflect a highly effective classifier, particularly for the labels Aggressive and Happy, with room for improvement in the classification of Romantic.

TABLE II
EVALUATION RESULTS OF THE RANDOM FOREST MODEL

Class	Precision	Recall	F1-Score
Aggressive	0.971	1	0.985
Dramatic	0.96	0.95	0.955
Happy	1	0.96	0.98
Romantic	0.876	0.92	0.898
Sad	0.938	0.91	0.924
accuracy	-	-	0.948
macro avg	0.949	0.948	0.948
weighted avg	0.949	0.948	0.948

C. *Support Vector Classifier Model*

Table 3 presents the results of solving a classification problem using the SVC (Support Vector Classifier). The SVC, a classification algorithm based on Support Vector Machine principles, was implemented in this experiment with a linear kernel to construct the model. The model was then trained and evaluated on a specific dataset. The precision for the 'aggressive' class was 93.4%, and the recall was 99%, indicating a high proportion of predictions correctly identified as 'aggressive.' The F1-Score, the harmonic mean of precision and recall, was recorded at 96.1%. Overall, the SVC results demonstrate that the model performs well generally and classifies the 'aggressive' class exceptionally well. However, the comparatively lower performance in the 'romantic' and 'ad' classes suggests that further model performance enhancement is necessary.

TABLE III
EVALUATION RESULTS OF THE SVC MODEL

Class	Precision	Recall	F1-Score
Aggressive	0.934	0.99	0.961
Dramatic	0.865	0.9	0.882
Happy	0.93	0.93	0.93
Romantic	0.862	0.81	0.835
Sad	0.854	0.82	0.837
accuracy	-	-	0.89
macro avg	0.889	0.89	0.889
weighted avg	0.889	0.89	0.889

D. *AdaBoost Model*

The AdaBoost classifier is a machine learning model that combines multiple "weak learners" to form a robust predictive model [41]. In this case, the AdaBoostClassifier from the sci-kit-learn library is used, which defaults to using the DecisionTreeClassifier as its weak learner. AdaBoost is particularly sensitive to noisy data and outliers, which might explain the lower performance in some classes. Table 4 shows a breakdown of the model and its performance metrics:

- The `aggressive` class has a relatively high precision and recall, resulting in a solid f1-score of 83%.
- The `dramatic` and `sad` classes have lower precision and recall values, indicating challenges in accurately predicting these classes.
- The `happy` class has decent precision but lower recall, suggesting the model is conservative in predicting this class and misses some actual cases.
- While the `romantic` class has lower precision, it has the highest recall, indicating that the model tends to over-predict this class.

The model's overall accuracy is 53.0%, with the macro and weighted averages for precision, recall, and f1-score hovering around 52.4% to 54.7%. These results imply that while the model performs well in the `aggressive` class, it struggles with the other categories to varying degrees, leading to moderate overall performance. AdaBoost is particularly sensitive to noisy data and outliers, which might explain the lower performance in some classes. Fine-tuning parameters like the number of estimators and learning rate, or even using a different base estimator, could improve the model's predictive accuracy.

TABLE IV
EVALUATION RESULTS OF THE ADABOOST MODEL

Class	Precision	Recall	F1-Score
Aggressive	0.934	0.99	0.961
Dramatic	0.865	0.9	0.882
Happy	0.93	0.93	0.93
Romantic	0.862	0.81	0.835
Sad	0.854	0.82	0.837
accuracy	-	-	0.89
macro avg	0.889	0.89	0.889
weighted avg	0.889	0.89	0.889

IV. CONCLUSION

This paper underscores the importance of background music in OTT services and its impact on user experience, proposing a system that aids all users, including those with hearing impairments, better understand the emotions conveyed through background music. The system combines audio signal processing techniques such as MFCC, RMS, and Mel Spectrogram with various machine learning algorithms—Logistic Regression, Random Forest, AdaBoost, and SVC—to analyze the emotional characteristics of background music and convert them into textual subtitles. The experimental results validate the utility of this technology, with the Random Forest algorithm showing the highest accuracy. This system can be utilized to improve the accessibility of emotional elements in a wide range of multimedia content, not just OTT services, enabling all users, especially those with hearing impairments, to have a deeper understanding and enjoyment of content through music, a non-verbal communication channel.

While the proposed system for emotional analysis of background music and conversion to textual subtitles has significantly enriched user experience in OTT services, further research is anticipated in several areas. First, as the perception of emotions in background music can vary across different cultures and genres, there is a need to improve the universality of the model through research that includes a variety of cultural backgrounds and musical genres. Second, refining and expanding the emotion classification system used in the current study is essential to develop a model capable of recognizing and expressing a more sophisticated and varied range of emotional states. Third, for actual application in OTT services, creating a system capable of analyzing background music and generating subtitles in real time is planned.

Acknowledgment

The Basic Science Research Program supported this research through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (RS-2023-00246191)

References

[1] K. S. Sontakke, “Trends in OTT Platforms Usage During COVID-19 Lockdown in India,” *Journal of Scientific Research*, vol. 65, no. 08, pp. 112–114, 2021, doi: 10.37398/jsr.2021.650823.

[2] Kim, Woo-Hyeon, et al. “Multi-Modal Deep Learning Based Metadata Extensions for Video Clipping”. *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 1, Feb. 2024, pp. 375-80, doi:10.18517/ijaseit.14.1.19047.

[3] Gangwar VP, Sudhagoni VS, Adepu N, Bellamkonda ST. Profiles and Preferences of OTT users in Indian Perspective. *European Journal of Molecular & Clinical Medicine*. vol. 7, no. 8, 2020.

[4] M. Yasen and S. Tedmori, “Movies Reviews Sentiment Analysis and Classification,” *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Apr. 2019, doi: 10.1109/jeeit.2019.8717422.

[5] J. Kim, C. Nam, and M. H. Ryu, “IPTV vs. emerging video services: Dilemma of telcos to upgrade the broadband,” Telecommunications Policy, vol. 44, no. 4, p. 101889, May 2020, doi:10.1016/j.telpol.2019.101889.

[6] M. S. Nordin et al., "Stress Detection based on TEO and MFCC speech features using Convolutional Neural Networks (CNN)," *2022 IEEE International Conference on Computing (ICOCO)*, Kota Kinabalu, Malaysia, 2022, pp. 84-89, doi:10.1109/ICOCO56118.2022.10031771.

[7] M. Selvaraj, R. Bhuvana and S. Padmaja, "Human speech emotion recognition", *Int. J. Eng. Technol*, vol. 8, no. 1, pp. 311-323, 2016.

[8] Z. Fu, G. Lu, K. M. Ting and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation," *in IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 303-319, April 2011, doi:10.1109/TMM.2010.2098858.

[9] V. Bansal, G. Pahwa and N. Kannan, "Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks," *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, Greater Noida, India, 2020, pp. 604-608, doi: 10.1109/gucon48875.2020.9231094.

[10] S. A. A. Qadri, T. S. Gunawan, M. Kartiwi, H. Mansor and T. M. Wani, "Speech Emotion Recognition Using Feature Fusion of TEO and MFCC on Multilingual Databases", *Lecture Notes in Electrical Engineering*, vol. 730, pp. 681-691, 2022.

[11] Q. Li et al., "MSP-MFCC: Energy-Efficient MFCC Feature Extraction Method With Mixed-Signal Processing Architecture for Wearable Speech Recognition Applications," *in IEEE Access*, vol. 8, pp. 48720-48730, 2020, doi: 10.1109/access.2020.2979799.

[12] S. Masood, J. S. Nayal and R. K. Jain, "Singer identification in Indian Hindi songs using MFCC and spectral features," *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, 2016, pp. 1-5, doi:10.1109/icpeices.2016.7853641.

[13] J. Dutta and D. Chanda, "Music Emotion Recognition in Assamese Songs using MFCC Features and MLP Classifier," *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2021, pp. 1-5, doi: 10.1109/conit51480.2021.9498345.

[14] K. L. Ong, C. P. Lee, H. S. Lim, K. M. Lim and A. Alqahtani, "Mel-MViTv2: Enhanced Speech Emotion Recognition With Mel Spectrogram and Improved Multiscale Vision Transformers," *in IEEE Access*, vol. 11, pp. 108571-108579, 2023, doi:10.1109/access.2023.3321122.

[15] S. D. Handy Permana and T. K. A. Rahman, "Improved Feature Extraction for Sound Recognition Using Combined Constant-Q Transform (CQT) and Mel Spectrogram for CNN Input," *2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, Karawang, Indonesia, 2023, pp. 185-190, doi:10.1109/icmeralda60125.2023.10458162.

[16] Y. Khasgiwala and J. Tailor, "Vision Transformer for Music Genre Classification using Mel-frequency Cepstrum Coefficient," *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, Kuala Lumpur, Malaysia, 2021, pp. 1-5, doi: 10.1109/gucon50781.2021.9573568.

[17] S. -H. Cho, Y. Park and J. Lee, "Effective Music Genre Classification using Late Fusion Convolutional Neural Network with Multiple Spectral Features," *2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, Yeosu, Korea, Republic of, 2022, pp. 1-4, doi: 10.1109/icce-asia57006.2022.9954732.

[18] G. Ulutas, G. Tahaoglu and B. Ustubioglu, "Forge Audio Detection Using Keypoint Features on Mel Spectrograms," *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*, Prague, Czech Republic, 2022, pp. 413-416, doi:10.1109/tsp55681.2022.9851327.

[19] W. B. Zulfikar, Y. A. Gerhana, A. Y. P. Almi, D. S. Maylawati and M. I. A. Amin, "Mood of Song Detection Using Mel Frequency Cepstral Coefficient and Convolutional Neural Network with Tuning Hyperparameter," *2023 11th International Conference on Cyber and IT Service Management (CITSM)*, Makassar, Indonesia, 2023, pp. 1-6, doi: 10.1109/citsm60085.2023.10455644.

[20] K. Wang, C. Qian and L. Zhang, "Machine learning music emotion recognition based on audio features," *2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Dalian, China, 2023, pp. 215-220, doi:10.1109/iciscae59047.2023.10392981.

[21] W. Wang, "CNN based music emotion recognition," *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, Hangzhou, China, 2021, pp. 190-195, doi:10.1109/icaice54393.2021.00044.

[22] Melinda, Melinda, et al. “Design and Implementation of Mobile Application for CNN-Based EEG Identification of Autism Spectrum Disorder”. *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 1, Feb. 2024, pp. 57-64, doi:10.18517/ijaseit.14.1.19676.

[23] Haque, Radiah, et al. “Classification Techniques Using Machine Learning for Graduate Student Employability Predictions”. *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 1, Feb. 2024, pp. 45-56, doi:10.18517/ijaseit.14.1.19549.

[24] S. Khade, S. Gite, S. D. Thepade, B. Pradhan, and A. Alamri, “Detection of Iris Presentation Attacks Using Hybridization of Discrete Cosine Transform and Haar Transform with Machine

Learning Classifiers and Ensembles," *IEEE Access*, vol. 9, pp. 169231-169249, 2021, doi: 10.1109/access.2021.3138455.

[25] M. H. Baffa, M. A. Miyim, and A. S. D. Dauda, "Machine learning for predicting students' employability," *UMYU Sci.*, vol. 2, no. 1, 2023, doi:10.56919/usci.2123_001.

[26] L. S. Hugo, "A comparison of machine learning models predicting student employment," *J. of Chemical Information and Modeling*, vol. 53, no. 9. 2018, [Online]. Available: http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1544127100472053.

[27] S. Islam, T. Akter, S. Zakir, S. Sabreen, and M. I. Hossain, "Autism Spectrum Disorder Detection in Toddlers for Early Diagnosis Using Machine Learning," *2020 IEEE Asia-Pacific Conf. Comput. Sci. Data Eng. CSDE 2020*, 2020, doi: 10.1109/csde50874.2020.9411531.

[28] M. A. Siddiqi and W. Pak, "An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection," *IEEE Access*, vol. 9, pp. 137494-137513, 2021, doi: 10.1109/access.2021.3118361.

[29] T. Le Minh, L. Van Tran, and S. V. T. Dao, "A Feature Selection Approach for Fall Detection Using Various Machine Learning Classifiers," *IEEE Access*, vol. 9, pp. 115895-115908, 2021, doi:10.1109/access.2021.3105581.

[30] B. Wang and J. Zhang, "Logistic Regression Analysis for LncRNA-Disease Association Prediction Based on Random Forest and Clinical Stage Data," *IEEE Access*, vol. 8, pp. 35004-35017, 2020, doi:10.1109/access.2020.2974624.

[31] A. Lucas, A. T. Williams, and P. Cabrales, "Prediction of Recovery from Severe Hemorrhagic Shock Using Logistic Regression," *IEEE J. Transl. Eng. Heal. Med.*, vol. 7, no. June, pp. 1-9, 2019, doi:10.1109/jtehm.2019.2924011.

[32] Z. Zhang and Y. Han, "Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier with an Advanced Machine Learning Approach," *IEEE Access*, vol. 8, pp. 44999-45008, 2020, doi: 10.1109/access.2020.2977962.

[33] J. C. Nwadiuto, S. Yoshino, H. Okuda, and T. Suzuki, "Variable Selection and Modeling of Drivers' Decision in Overtaking Behavior Based on Logistic Regression Model with Gazing Information," *IEEE Access*, vol. 9, pp. 127672-127684, 2021, doi:10.1109/access.2021.3111753.

[34] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny).*, vol. 507, pp. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.

[35] Susetyoko, Ronny, et al. "An Improved Accuracy of Multiclass Random Forest Classifier With Continuous Attribute Transformation Using Random Percentile Generation". *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, no. 3, June 2023, pp. 943-5, doi:10.18517/ijaseit.13.3.18379.

[36] R. Susetyoko, W. Yuwono, E. Purwantini, and B. N. Iman, "Characteristics of Accuracy Function on Multiclass Classification Based on Best, Average, and Worst (BAW) Subset of Random Forest Model," pp. 410-417, 2022, doi: 10.1109/ies55876.2022.9888374.

[37] M. A. Ganaie, M. Tanveer, P. N. Suganthan, and V. Snasel, "Oblique and rotation double random forest," *Neural Networks*, vol. 153, pp. 496-517, 2022, doi: 10.1016/j.neunet.2022.06.012.

[38] M. Gencturk, A. Anil Sinaci, and N. K. Cicekli, "BOFRF: A Novel Boosting-based Federated Random Forest Algorithm on Horizontally Partitioned Data," *IEEE Access*, vol. 10, no. August, pp. 89835-89851, 2022, doi: 10.1109/access.2022.3202008.

[39] C. Zou et al., "Heartbeat Classification by Random Forest With a Novel Context Feature: A Segment Label," *IEEE J. Transl. Eng. Heal. Med.*, vol. 10, no. August 2022, doi: 10.1109/jtehm.2022.3202749.

[40] D. A. Anggoro and N. A. Afdallah, "Grid Search CV Implementation in Random Forest Algorithm to Improve Accuracy of Breast Cancer Data," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 12, no. 2, p. 515, Apr. 2022, doi:10.18517/ijaseit.12.2.15487.

[41] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE Access*, vol. 9, pp. 165286-165294, 2021, doi:10.1109/access.2021.3134330.