electronics

MDPI

Article

Classical Music Specific Mood Automatic Recognition Model Proposal

Suyeon Lee [1], Haemin Jeong [2] and Hyeyoung Ko [2,*]

[1] Department of Computer Science & Engineering, Seoul Women's University, Seoul 01797, Korea; syou93@swu.ac.kr
[2] Department of Digital Media Design and Applications, Seoul Women's University, Seoul 01797, Korea; wjdgoalss@swu.ac.kr
* Correspondence: kohy@swu.ac.kr; Tel.: +82-2-970-5751

Abstract: The purpose of this study was to propose an effective model for recognizing the detailed mood of classical music. First, in this study, the subject classical music was segmented via MFCC analysis by tone, which is one of the acoustic features. Short segments of 5 s or under, which are not easy to use in mood recognition or service, were merged with the preceding or rear segment using an algorithm. In addition, 18 adjective classes that can be used as representative moods of classical music were defined. Finally, after analyzing 19 kinds of acoustic features of classical music segments using XGBoost, a model was proposed that can automatically recognize the music mood through learning. The XGBoost algorithm that is proposed in this study, which uses the automatic music segmentation method according to the characteristics of tone and mood using acoustic features, was evaluated and shown to improve the performance of mood recognition. The result of this study will be used as a basis for the production of an affect convergence platform service where the mood is fused with similar visual media when listening to classical music by recognizing the mood of the detailed section.

Keywords: classical music segmentation; classical music mood; mood recognition model; machine learning; emotional intelligence

Citation: Lee, S.; Jeong, H.; Ko, H. Classical Music Specific Mood Automatic Recognition Model Proposal. *Electronics* **2021**, *10*, 2489. https://doi.org/10.3390/electronics10202489

Academic Editors: Manohar Das and Xianzhi Wang

Received: 3 September 2021
Accepted: 11 October 2021
Published: 13 October 2021

1. Introduction

Music of various moods has been shown to enrich the listener's emotions and promote psychological and mental stability [1–5]. Classical music is longer than other music genres, so it contains various detailed mood changes in one song [6]. If a specific classical music mood can be recognized accurately, it can be applied to various objectives, such as expanding from a single appreciation of music and fusing with visual media of a similar mood to expand the emotional experience.

The recent development of music streaming services, such as YouTube and Spotify, is linked to the rapid development of music information retrieval (MIR) technology which efficiently retrieves the required information from among vast amounts of music data. In MIR, in addition to the basic metadata of music, many technologies for pattern recognition, clustering, extraction automation, and the advancement of acoustic features are being developed through the analysis of music signals [7–10].

In addition, many studies are being performed in the field of music emotion recognition (MER) to classify or recommend music based on emotion. MER recognizes emotions for the whole or segmented music. In the case of music with lyrics, there is also a study addressing the recognition of emotions by interpreting the lyrics [11,12]. However, as there are cases in which there are no lyrics or the structural characteristics of music are more related to emotions rather than the lyrics, many studies are being conducted to recognize the emotions of music by analyzing the acoustic features of the music. In addition, various

studies related to the service application fields, tailored to the sensibility of music, are being conducted [13–16].

Conventional MER studies for the purpose of searching for the emotion of music, were mainly conducted to grasp the overall emotion of the music rather than the emotion for each sub-section of the music. On the other hand, Wu et al. (2014) [17] and Xiao et al. (2008) [18] conducted a study on the sensibility of sub-sections of music in order to understand the mood of the music. This is because, in the case of long music, it is possible to clearly derive the mood by understanding the emotions of the detailed sections of the music rather than understanding the overall emotions.

In previous research for recognizing the flow of a detailed music mood, the music segmentation was comprehensively executed in similar intervals and the study focused on a mood analysis based on this type of segmentation [6,19]. However, in order to clearly distinguish the units for recognizing the detailed mood of the music, studies on recognizing the mood after extracting the sections of similar acoustic features are also being conducted [20,21].

The direction of this research study is to create the capability of combining the specific detailed aesthetics of mood that constitute classical music and apply them to other media which contain similar moods. Therefore, in this study, an effective model is proposed that can automatically recognize the detailed mood of classical music, so that a foundation can be established per the direction of the study. To accomplish this, it is necessary to increase the specificity of the detailed mood recognition in classical music, and the specificity must satisfy the following detailed conditions. The segmentation of the music segment, which is the analytical unit of the detailed mood, must be reasonable. The performance of the segment's mood analysis must be high, and the music segment unit and the mood analysis result must have appropriate criteria to ensure that a service can have the ability to combine with various future media.

In order to derive results that satisfy these conditions, this study intends to be conducted in the following steps. Section 2 establishes a theoretical foundation based on previous studies. After the theoretical assessment, Section 3 involves activities in music data segmentation, model learning among music segments, and finally the execution of mood training data labeling by a music expert. Based on the data from these activities, we create a mood recognition model (MRM) of classical music. In Section 4, the performance of the music segment extraction method in the proposed MRM in this study and the validity of the algorithm selected for use in the model are evaluated. Finally, Section 5 provides conclusions to the study as well as directions for future research.

2. Related Works

2.1. Emotional Representation

Researches in psychology on feeling and the inner state of humans are representatively explained by the basic emotional model and affective dimensional model. First, James [22] and Ekman [23], who researched the basic emotional model, defined individual emotions as independent beings between emotions. A basic emotional model by Ekman, which is widely used in various fields, defined emotions as anger, disgust, fear, enjoyment, sadness, and surprise. Moreover, emotions form unique features as humans have evolved and each emotion has apparent features of an independent signal, physiologic expression, and antecedent event. In particular, interculturally featured basic feelings which are apparent signals apart from other feelings are shown by facial expressions. Therefore, basic emotions of the basic emotional model have obvious features that are useful in general apparent feeling recognitions and expression fields [24–31].

Russell [32] is a representative researcher of the dimensional model related to "Affect". Russell suggested an affective dimensional model with affect adjectives in 6–12 bipolar circular dimensional arrangements. The researcher established pleasure–misery as a horizontal dimension and arousal–sleepiness as a perpendicular dimension. He then selected 28 words describing mood, feeling, affect, and emotion as representative affect adjectives,

and suggested a 2-dimensional model in a circular arrangement. This affective dimensional model is used for classification and interpretation based on the axes and thereby trending an inner state.

Yang et al. (2008) [33] used the values of 'Arousal' and 'Valence' of music samples to conduct research for music emotion recognition (MER) by using the dimensional model. The research allowed effective browsing of music samples when the subject selected a specific coordinate value on the Arousal-Valence plane with Arousal-Valence (AV) as axes. The Arousal-Valence Affective Dimensional Model used a regression approach to defining feelings by AV values to improve the ambiguity of the categorization of conventional feelings.

Thayer (1989) [34] is another representative researcher of the Dimensional Model who researched a simplified and organized Dimensional Model by using mood words to improve overlapping and ambiguous mood adjectives from the Russell model. Thayer referred to Russell's Affective Model to define a 2-dimensional mood by establishing 'Stress' and 'Energy' as categories to dictate the extent of pleasure to displeasure by plotting the extent of 'Stress' and the extent of calm to energetic by plotting the extent of 'Energy.' In addition, moods are divided into quadrants of 'Exuberance', 'Contentment', 'Anxious', and 'Depression' based on the corresponding plotted coordinates on this 2-dimensional plane. If Thayer's Mood System is applied to music, 'Energy' equates to the strength and intensity of the music and 'Stress' equates to the tone and tempo of the music. The advantage to this system was the capability of distinguishing mood adjectives by categorization.

On one hand, Lu et al. (2006) [6] emphasized that Thayer's mood adjectives were rational in expressing the mood of the music and therefore studied classifying moods of music by using these adjectives. However, Yang et al. (2007) [35] proposed in a study on group-wise music emotion recognition (GWMER) and personalized music emotion recognition (PMER) that Thayer's Mood system lacked numbers of mood adjectives that would subjectively express an individual's personal preference and identity as shown in Figure 1.

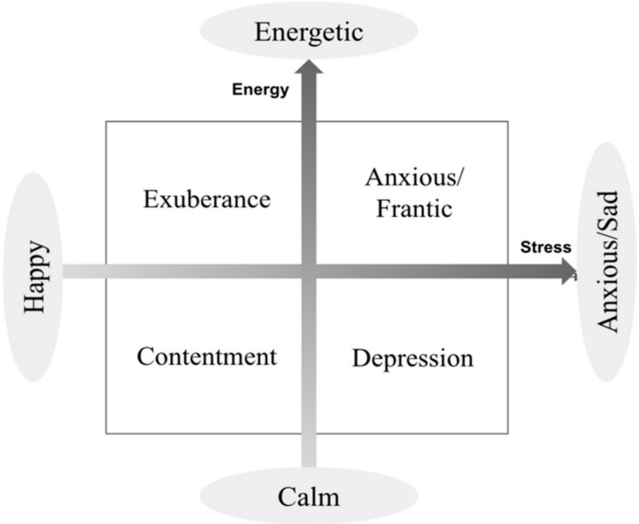

Figure 1. Two-Dimensional Structure of Thayer's Mood System.

Meanwhile, in various psychological studies, terms such as 'mood', 'emotion' and 'affect' are used to describe human feeling or inner state. These words contain similar, or related concepts. 'Affect' is used as a word to include emotion and mood. Emotion and affect have a shorter duration than mood and are immediately revealed by facial expressions, whereas mood has a long duration and is void of expression. Emotion and affect do not affect changes in human behavior or nature, but mood is considered to have a potential impact on them [23]. Mood is generally used quite frequently as a term related to human feeling through music. This is because, rather than the characteristics expressed through facial expressions in a short period of time, it is interpreted in conjunction with the

characteristics of music that can affect the inner state of human beings for a longer period of time and ultimately affect the nature or behavior in the long run.

This research intends to use mood adjectives for a study related to the feeling of classical music. In addition to defining words with clear individual characteristics between adjectives, which is the strength of the individual emotion model in the existing field of psychology, this study intends to increase the number of representative adjectives to effectively represent the feeling of the music. We additionally intend to define the mood by using a dimensional model that is good for explaining the organic relationship between words. First, in order to collect various mood adjectives in music, we intend to use the mood tag [36], which is used in the music service 'All music'. 'All music' provides users with music albums labeled by music experts with 289 different mood adjectives related to music. Music mood adjectives used in the service can be used as data for selecting adjectives specialized in classical music in this study. However, since the number of adjectives with similar meanings is large and the range of adjectives to express various music genres is vast, clustering and selection processes of adjectives are necessary to utilize them in this study.

2.2. Acoustic Features

In general, music has unique acoustic features such as rhythm, harmony, pitch, tone, and tempo. The acoustic feature can be derived through the analysis of audio signals, which are being dealt with through various studies such as clustering similar sections of music or using them as a unit to search for music. For the definition of acoustic features, Tzanetakis and Cook (2002) [37] classified classical music into three major categories, namely timbral texture features, rhythmic content features, and pitch content features.

Weihs et al. (2007) [38] divided the acoustic features into short-term features and long-term features according to the length of the music section to be extracted. Among the acoustic features, features such as a specific tempo and instrument, which are intuitively easy to explain, were separately classified as semantic features. Scaringella et al. (2006) [39] defined three groups of acoustic features related to the major dimensions of music: timbre, rhythm, and pitch information.

On the other hand, Fu et al. (2010) [40] took the acoustic feature from the work of Weihs et al. (2007) [38] and Scaringella et al. (2006) [39] and integrated the taxonomy of those two studies to hierarchically classify and define them, as shown in Figure 2. The low-level feature consists of timbre features that capture tone and temporal features that capture the change and evolution of tone over time. The low-level feature is composed of various other features in detail, and the extraction performance is excellent even with a simple procedure, so it is mainly used for recognizing the mood, genre, and instruments of music. The mid-level feature is a feature that the listener can recognize, and the acoustic features are classified systematically in an integrated way by including mainly three features: rhythm, pitch, and harmony.

2.3. Audio Signal Analysis & Feature Extraction

The increase in users' needs for music content and the rapid growth of music-related streaming service markets, such as YouTube and Spotify, are closely related to the advancement of the MIR field. In order to efficiently perform MIR, it is necessary to analyze audio signals that can automate and advance the extraction of acoustic features from music data. For this purpose, various acoustic feature extraction frameworks, such as Marsyas [7], PsySound3 [8], MIRtoolbox [9], and LibROSA [10], have been proposed. These frameworks recognize and analyze audio signals, extracts acoustic features, such as pitch, harmonicity, spectral centroid, spectral moments, mel-frequency cepstral coefficients (MFCC), and analyzes acoustic features to provide pattern recognition or clustering function. The LibROSA [10] framework was developed under a python environment with good expansion by being optimized to big data and machine learning, deep learning, and other related environments. LibROSA [10] uses short time Fourier transform (STFT) to extract audio

features, such as chroma, melspectrogram, MFCC, RMS, and tempogram, from audio signals. An algorithm is provided by analyzing this extracted acoustic features to perform clustering between the same intervals with the same characteristics. Bayu et al. (2019) [41] used LibROSA to extract acoustic features to create a model for classifying music emotions. And in Babu et al. (2021) [42]'s Speech Emotion Recognition System construction study, LibROSA was used by utilizing the LibROSA Library in a Python environment. In this study, when receiving new classical music data, we are also building a model that automatically recognizes the music mood for a detailed section in real time. In order for the research derived model to be utilized as various AI-based application services, it was considered to be efficient to build a model in a Python environment. Therefore, in this study, the structure of music based on acoustic features of classical music data is analyzed and the LibROSA Framework is used for mood recognition.

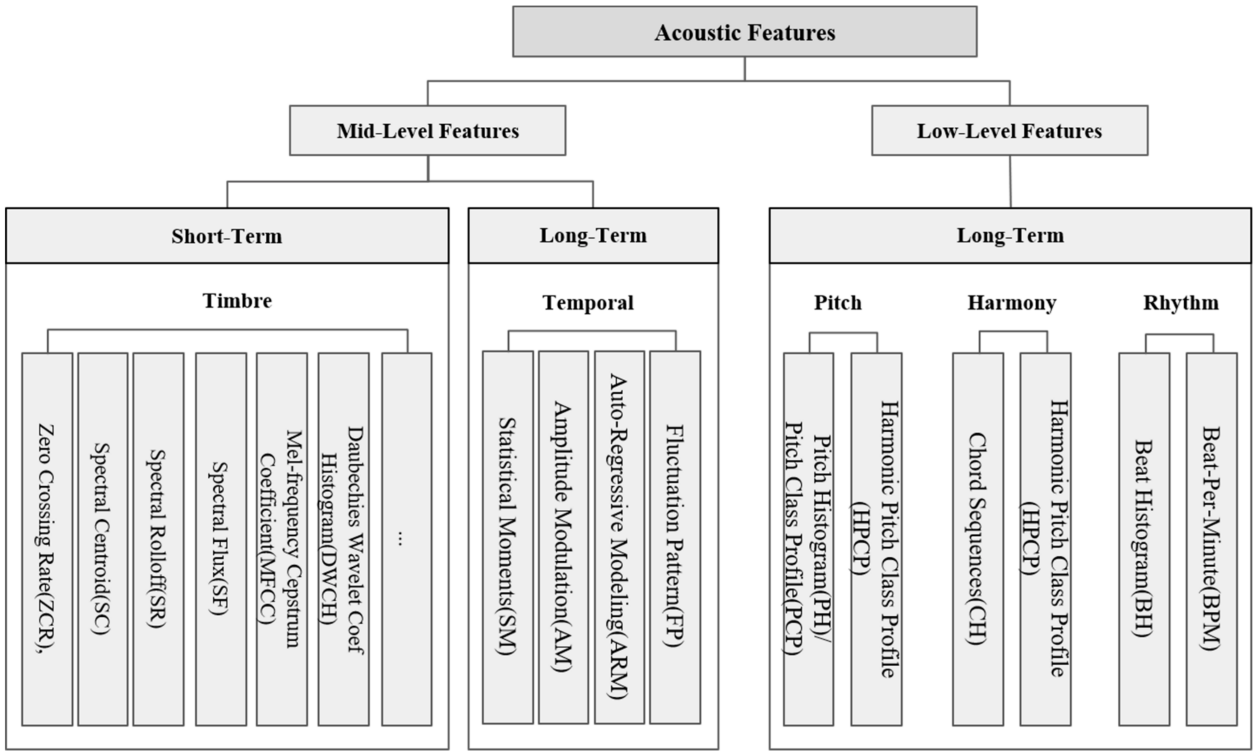

Figure 2. Hierarchical Classification of Acoustic Features [40].

2.4. Music Emotion Recognition

Among the previous studies on MER, many studies have been conducted to recognize and classify the emotion to a piece of music. However, the emotion of music has a time-varying characteristic that is dependent on the flow of playback time. In particular, as the length of music becomes longer, as in classical music, recognizing and classifying the whole music as a single emotion may have low classification accuracy and may be ambiguous. In addition, in order to link music to various services or media, there may be a limit to searching for detailed music emotions. Therefore, in this study, in order to recognize the detailed emotions of classical music as mood adjectives, it is necessary to study the mood emotion recognition method for each segment based on the segment extraction of music.

Among previous studies on recognizing music segments, Li and Ogihara et al. (2006) [13] and Han et al. (2010) [14] defined mood adjectives. First, Li and Ogihara et al. (2006) [13] divided classical music into several segments, extracted the timbre features of each segment, and conducted a study on mood recognition.

In this study, using the SVM model, three bipolar adjectives were defined as mood adjectives to be used in the study, mood was predicted by binary classification, and an

accuracy of about 70% was achieved. Moreover, Han et al. (2010) [14] conducted a study on recommending suitable music to evoke the user's desired emotional state. On the other hand, although classification using SVM can classify only one class at a time, SVM is used because it has the strength of SVM performance and simplicity of expression of verification results.

Lee et al. (2010) [15] conducted a study to discriminate the mood of each segment to recommend music suitable for the user's situation, after segmenting a section with structurally similar acoustic features in music. For mood recognition by segment, the individual subjective mood was modeled by using Thayer's Mood System through a regression analysis method.

Seo et al. (2019) [16] also used Thayer's Mood System. This study proposed a method of classifying music moods based on the Thayer's Mood System to automatically recommend music according to people's emotions in music-related applications. In the aforementioned studies, acoustic features were extracted for each segment of music and used as training data for a model predicting mood for each segment of music.

In this study, a total of 19 acoustic features, including rhythm, harmony, pitch, and timbre, among the acoustic features of classical music are extracted based on previous studies related to MER. Based on this, we intend to create a model that recognizes mood. For a model that recognizes the music mood, we intend to use XGBoost, which has a better performance than SVM used in the existing MER studies.

3. Method

In this study, we are going to study a model that automatically recognizes the detailed mood of classical music so that the detailed mood can be used for application services by fusing them to media of various moods. For the purposes of studying the model, the study is executed according to the study method and transitions shown in Figure 3. First, classical music and emotional adjective data are prepared. Subsequently, a long classical music data segmentation is created. Then, when the mood training data labeling of the music segments is completed, the Mood Classification Model learning is performed.

To prepare for the classical music research, 12,305 classical music data in total were collected from 5 organizations, such as FMA (Free Music Archive) [43], Musopen [44], Musicnet [45], KkachilhanClassic [46], and URMP [47]. In this study, music data must undergo segmentation processing, so the music data collection criteria were defined as audio data that can be used commercially, CC0 1.0 Universal, which is a license for secondary processing, and Public Domain Mark 1.0.

3.1. Music Data Segmentation

3.1.1. Acoustic Features Extraction and Music Data Segmentation

Classical music is composed of various acoustic features according to the flow of playing time, so that various detailed moods can be defined. For the unit to analyze the detailed mood of classical music, it is necessary to analyze the mood of the segment of classical music. There are two methods of segmenting classical music. One is the segmentation of the flow of music at regular time intervals. The other is the automatic segmentation of sections with similar acoustic features which can be a factor in affecting mood changes in music. A music Dataset A segmented at regular intervals of 30 s a music Dataset B segmented for structural characteristics of music are generated.

In general, classical music has a long playback time and includes countless audio spectrum analysis data, so an efficient method for analyzing the structure and segmentation is needed. In addition, effective acoustic features are required for segmentation according to the structural characteristics of music. In the case of MFCC, it is an audio feature that well represents the tone of music and is defined as the main acoustic feature of the emotional classification and genre classification models. In this study, to prepare dataset B, we use the Librosa framework to facilitate the extraction of music segment data.

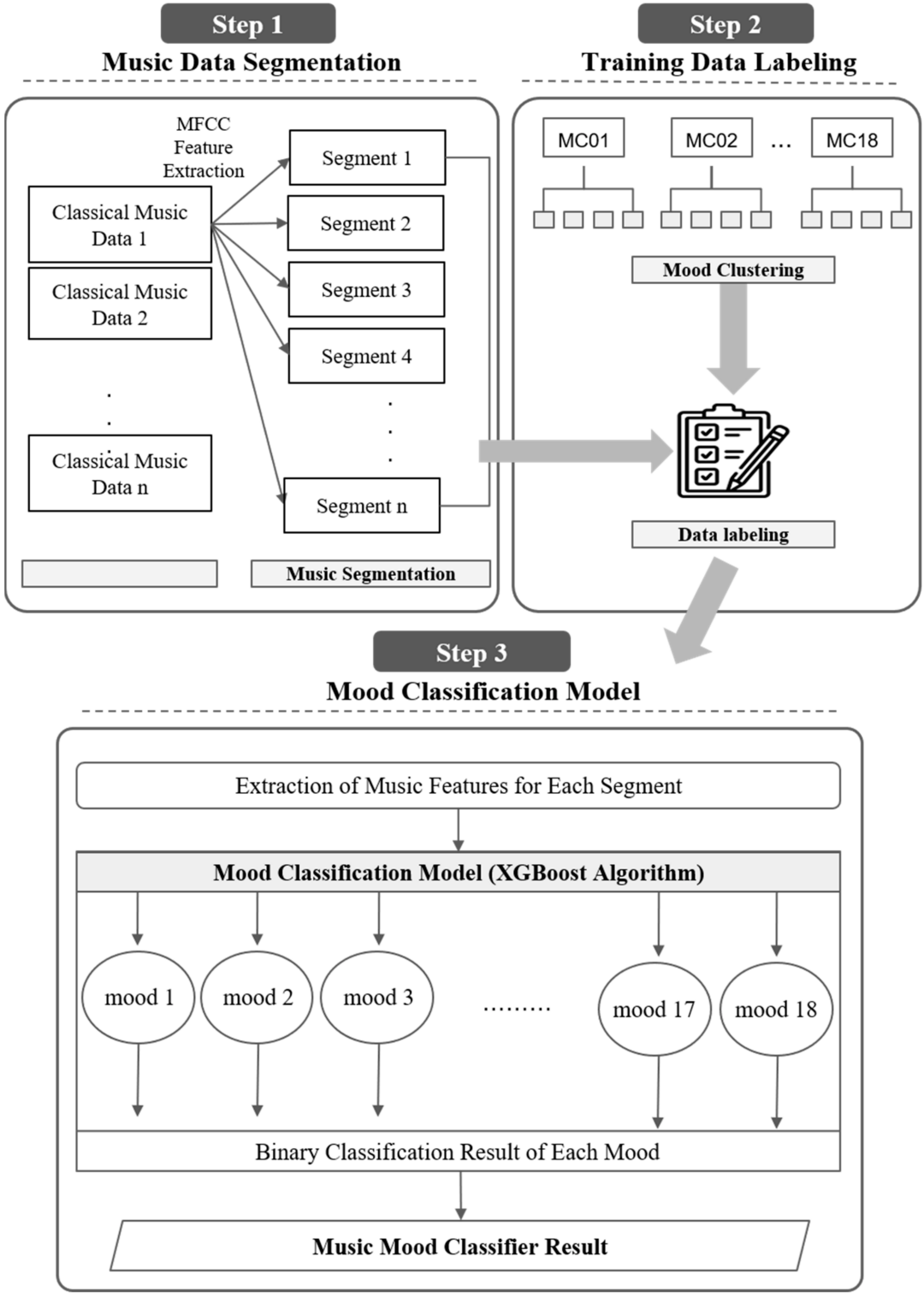

Figure 3. Introduction of Research Method per Stage.

MFCC uses a logarithmic scale at high frequencies to reflect the characteristics of human hearing organs by performing fine spectral decomposition at low frequencies in a linear low frequency subband [40]. After extracting these MFCC and by using the difference of features between previous frames, audio signals go through agglomerative clustering and audio signals with a large difference between feature values are determined to be the point of change, so data segmentation is processed.

Agglomerative clustering is one of the hierarchical clustering algorithm methods that assigns each data point as one cluster and combines two similar clusters until a designated number of clusters remains. In other words, the clustering method can detect a point where MFCC characteristic changes by the segment.agglomerative function. Moreover, the agglomerative clustering algorithm divides an input which is used to define designated k numbers of continuous components. In this study, since one piece of classical music should be divided into k segments, k was determined as the quotient obtained by dividing the length of classical music by 10. Therefore, segments are produced as a quotient of the length of classical music divided by 10 from one piece of classical music. The reason for the division by a factor of 10 is because it was considered appropriate to show visual media for one hour divided into 10 equal parts, considering that visual media with a mood similar to the music segment and visual media with a similar mood can be combined and serviced later. However, if all music is divided into 10 sections, a very short section can be created. We attempted to solve this case through the music segment clustering algorithm, as shown in Figure 4.

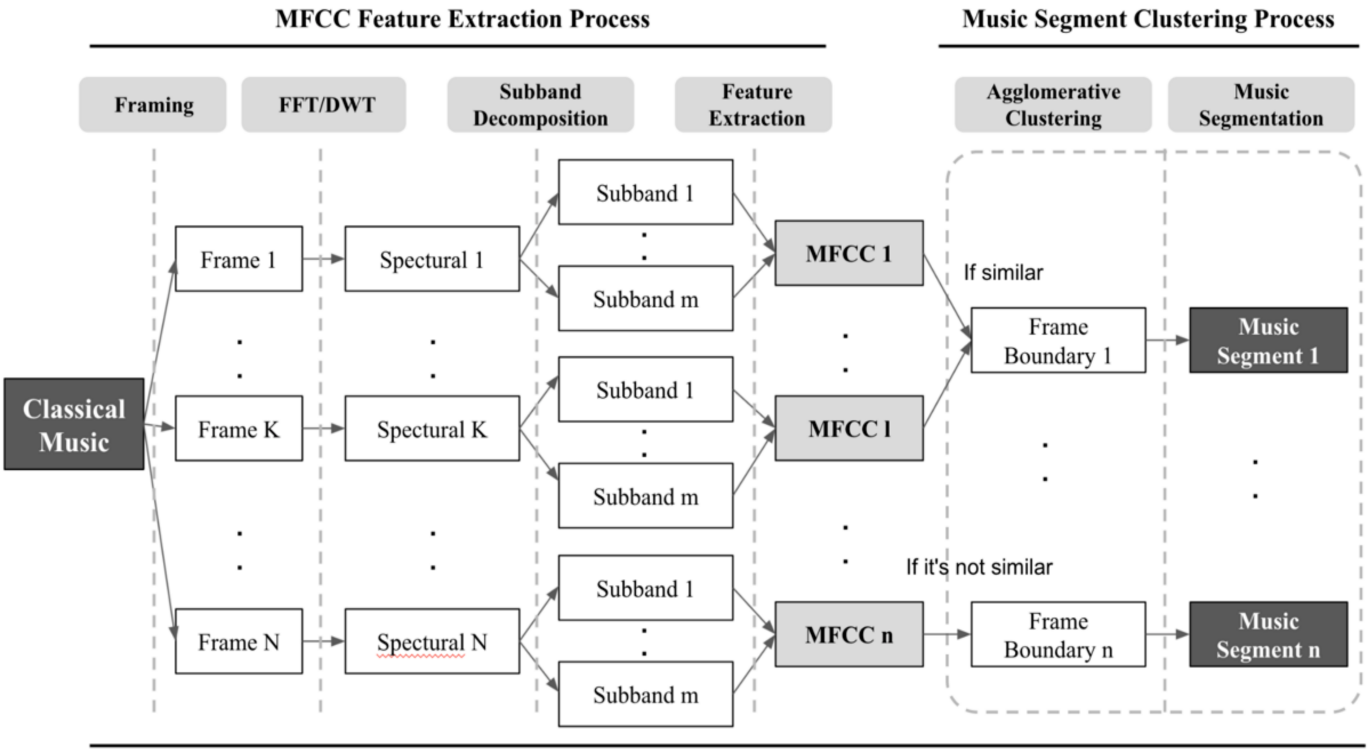

Figure 4. Dataset B Production Process by MFCC Feature Extraction and Music Segment Clustering.

Using a total of 12,305 classical music data collected for this study, Dataset A is segmented at regular intervals of 30 s and Music Dataset B is segmented according to the structural characteristics of music.

1. Dataset A Segements created by regularly dividing 30 s (data size: 1644 pieces).
2. Dataset B Segments created by dividing according to the proposed method (data size: 12,305 pieces).

Figure 5 shows Edward Elgar's 'Pomp and Circumstance—March No. 1 in D major op. 39-1′. The example shows a part of Dataset A in which 39-1′ is uniformly divided for 30 s and a part of Dataset B in which the segment is created by clustering between sections in which the MFCC features of the method proposed in this paper are similar. In the case of Edward Elgar's 'Pomp and Circumstance—March No. 1 in D major op. 39-1′, although it is the same classical music, the number of segments generated was 9 when performing

the segment using the Dataset A method, and 13 music segments were generated when performing the segmentation using the Dataset B method.

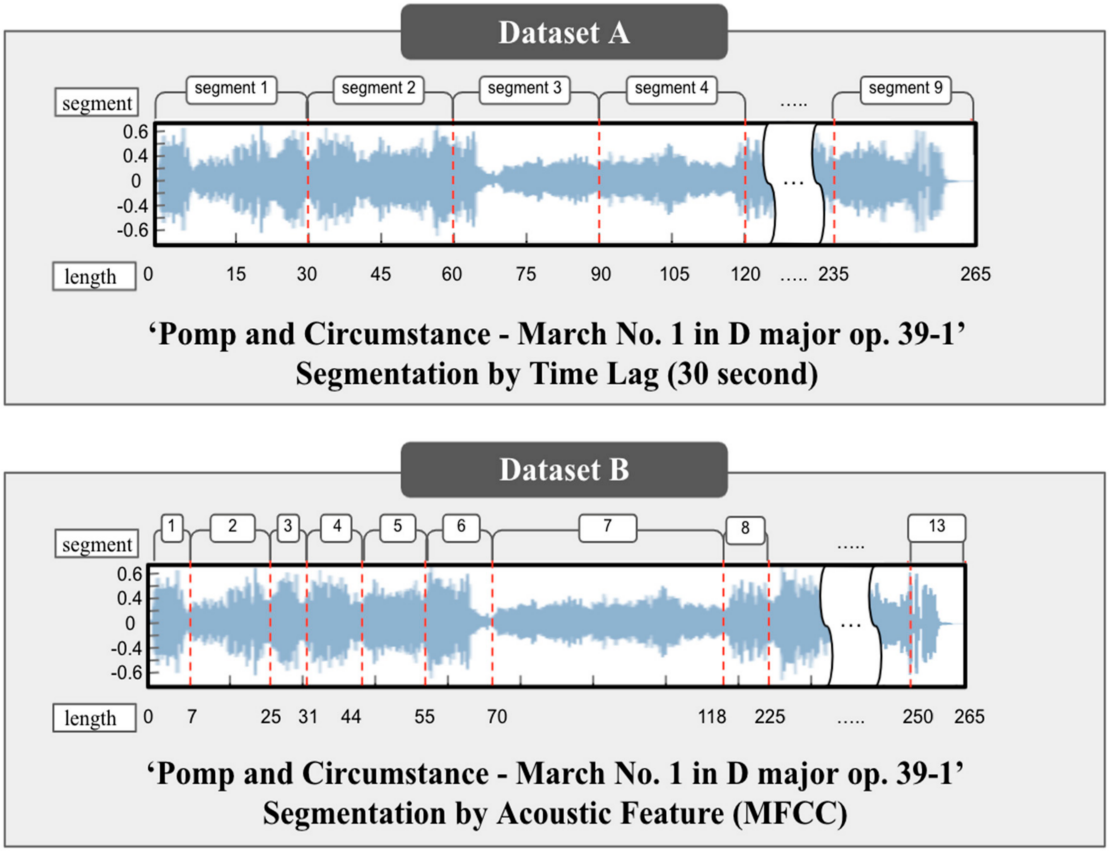

Figure 5. Music Data Segmentation Overview of Dataset A and Dataset B.

3.1.2. Clustering Algorithm

When music data is segmented by the segment extraction method using MFCC to produce dataset B, in the case of a song performed by an actual classical music performance, a segment of 5 s or under may be created. An example of such a short segment would be data that only contain applauding sounds. Since each segment must be an affectively meaningful interval, it is necessary to process a very short interval. This research added music segment short term clustering algorithm to cluster data 5 s or under to other clusters referring to the research of Xiao et al. (2008) [18], who suggested the length of aa music segment in which the mood of the music feels stable is 8 s and 16 s. In this study, the music segment short term clustering algorithm is added based on the fact that Xiao et al. (2008) [18] found that the stable music segment lengths that can effectively convey the music atmosphere are 8 s and 16 s.

The music segment short term clustering algorithm is an algorithm for merging music segments with other segments if the length of the music segment is 5 s or under. Since the first segment has no previous section, it is merged with the next segment. As a result, among all segments of the music, music segments of 5 s or under are merged with the previous or next segment to have a length of more than 5 s. By adding the music segment short term clustering algorithm, it is possible to stably match the mood for each music segment. Figure 6 below shows the process of clustering segments of 5 s or under using the music segment short term clustering algorithm in addition to the method of automatically segmenting music according to the MFCC feature when creating Dataset B.

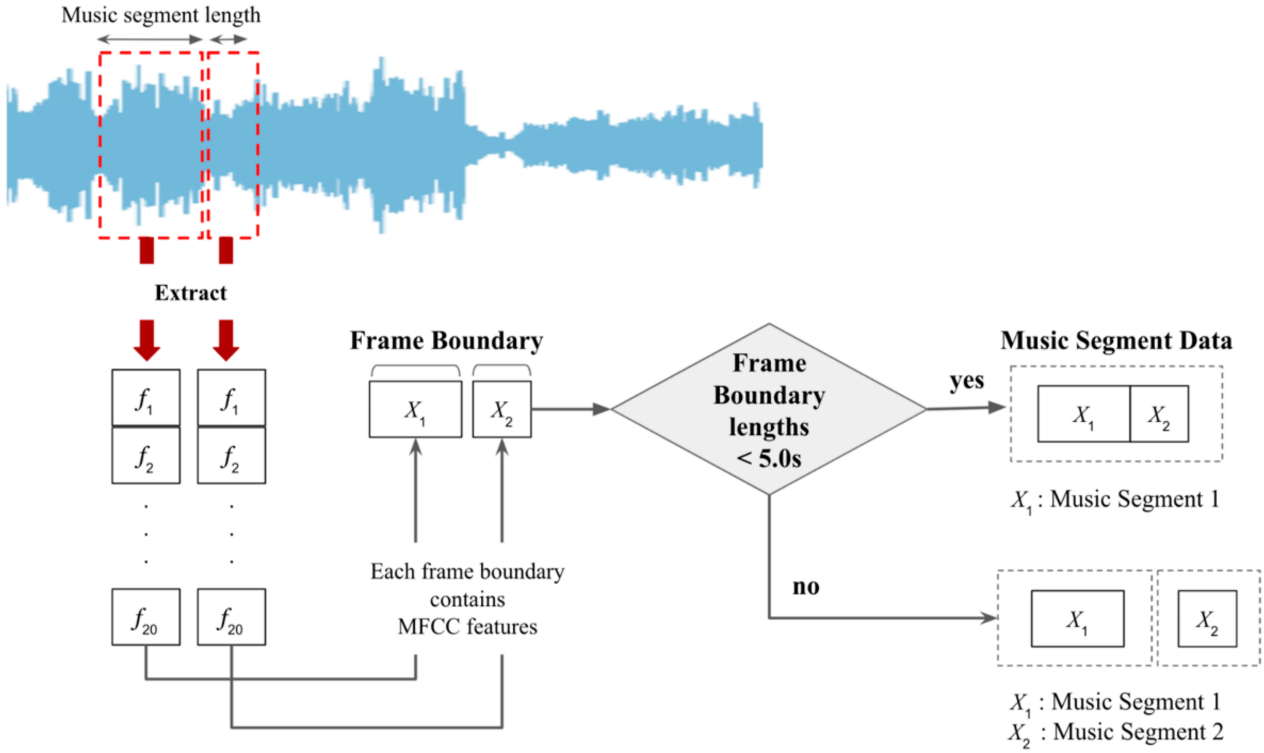

Figure 6. Overview of Music Segment Short Term Clustering Algorithm.

3.2. Training Data Mood Labelling

For the emotional labeling of the segmented classical music training data, the mood adjective class definition should be first established. This is to define the final class by performing the mood adjective clustering task through the affect, music expert group. By using mood adjectives defined on music training data which are each music segments in Dataset A and Dataset B by classical music panels, labelling is performed. As such, we intend to use the music labeling data generated through classical music experts as training data for the mood classification model of the music segment. Meanwhile, the mood labelling work of music segment data is performed by experts not by laymen to reduce the bias of data and provide clarity for mood extraction.

3.2.1. Definition of Classical Music Mood Adjective Class

In order to label the mood adjectives for each segment of the classical music segment data Dataset A and Dataset B prepared earlier, first, it is necessary to define an adjective class that reflects the characteristics of classical music well and considers the linkage of the affect convergence service of classical music in the future. There is a basic emotion model that proposes Ekman's basic emotional adjectives (anger, disgust, fear, enjoyment, sadness, surprise), the strength of which is to be clearly distinguished based on general emotions. For this model, there may be less music data collection for certain emotional adjective classes, such as disgust. Moreover, since it is an adjective that is faithful to the basic emotion connected to the expression, there are few options to reveal the mood in classical music, so there may be a limit to labeling the music mood. There is a need for an adjective class that can broaden the choice of affect adjectives, minimize overlapping or ambiguous adjectives, and effectively label the mood of music.

For this reason, this research used 289 mood tag adjectives used in music services, selected by music experts in 'All Music' which is an actual music streaming service to collect mood adjectives suggested for classical music mood. Among the 289 musical mood adjectives collected from 'All Music', similar mood adjectives are clustered except for those not related to classical music. Through this, the division between adjectives is reduced, but

the choice of adjectives is wider than the basic emotional adjectives, and it can be used efficiently for a mood suitable for listening to and expressing classical music.

For the definition of mood adjectives, two affect research experts used 289 mood adjective cards, referring to Russell's [32] affect model study correlation analysis results, Collins English Thesaurus [48], and Oxford Thesaurus [49]. Based on the thesaurus, 68 mood adjectives were first clustered. In this process. In this process, 68 adjectives are derived by excluding those related to hip-hop, drugs, and profanity that do not fit the classical music genre. And 68 adjectives were secondarily clustered into 19 adjectives. Figure 7 shows the clustering method of classical music mood adjectives and the finally drawn classical music mood adjective class.

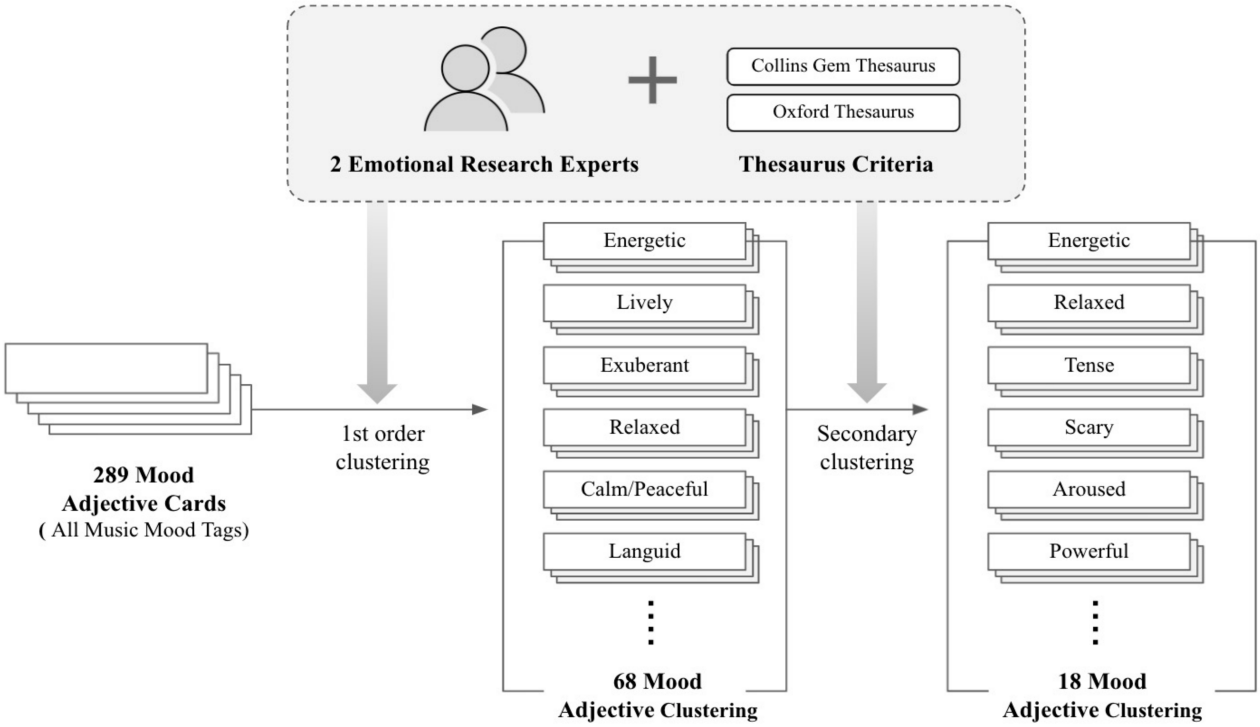

Figure 7. Classical Music Mood Adjective Clustering process.

3.2.2. Music Expert Data Labeling Method

In order to aid a model in learning to classify the mood of the classical music segment, we proceed with mood adjective labeling on the segment data. In this research, 3 classical music major experts labelled 1644 Dataset A segments and 12,305 Dataset B segments. After randomly selecting each interval of segment data per each classical music song from Dataset A and Dataset B that were prepared to verify a suitable music as training data, the labelled data are used as a training data for the mood classification model of the music segment.

The labelling of Dataset A and Dataset B is performed via the same method. The mix of Dataset A and Dataset B was handed over to be labelled by 3 music experts on 25 February 2021 and the experts were allowed to label multiple moods felt from the music interval using 18 mood adjectives selected for classical music. By allowing all moods to be selected after listening to a section of music, it was possible to classify the emotions of a section of music into a complex mood rather than a single mood. The music experts completed labelling all 13,949 music segments in total for 4 months from 25 February 2021 to 7 July 2021.

Table 1 shows a part of the results labelled by music experts on music segment data. The results include the music title, labelled segment number, start second and end second of the segment, segment length, and labelling information of segment.

Table 1. Classical Music Segment Data Labelling Result Paper of Classical Music Experts (part).

Music Title	Seg	Time (Start Second, End Second)	Length	Emotion			
				Energetic	Powerful	Joyous	Scary
Cello Sonata in G minor op. 65: II. Scherzo—Allegro con brio	9	[105.5, 113.5]	8		☑		☑
12 Etudes op. 10: No. 2 in A minor, Allegro	3	[45, 50.2]	5.2	☑	☑		
Nocturnes: No. 5 in F sharp major op. 15-2	14	[136.5, 152.6]	16.1				
Symphony No. 1 in C major op. 21: III. Menuetto. Allegro molto e vivace	16	[193.4, 209.5]	16.1	☑	☑	☑	

3.3. Classical Music Mood Recognition Model Production

In order to automatically recognize the various detailed moods that change in classical music, classical music has to be segmented. Based on this, a model that can automatically perform mood recognition for data is produced, as shown in Figure 8. In order for the classical music mood recognition model to learn mood recognition, the following processes must be conducted. First, acoustic features are extracted, which will be used for analyzing mood recognition from segment data. Next, a classical music mood recognition model is produced using the XGBoost algorithm.

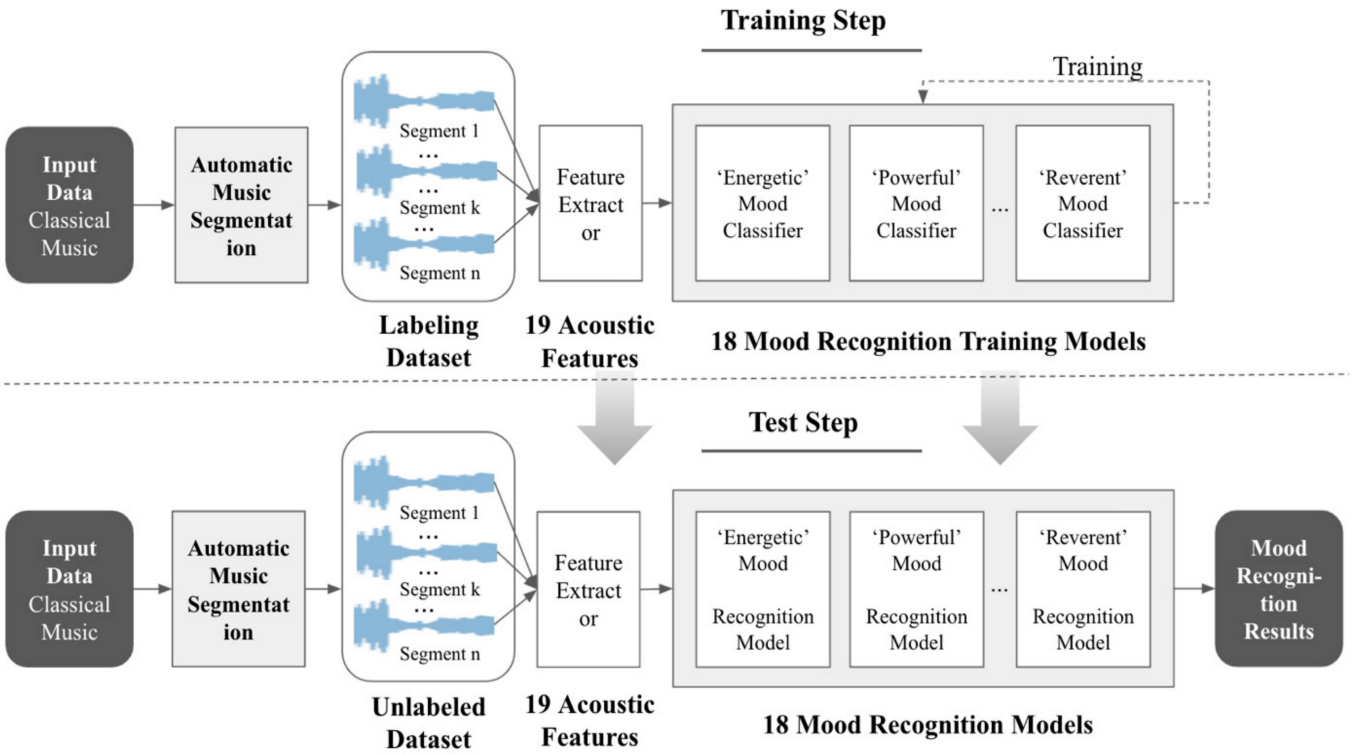

Figure 8. Classical Music Mood Recognition Model.

3.3.1. Extraction of 19 Acoustic Features for Mood Analysis

First, the acoustic features to be used for signal analysis for mood recognition of segment data are by using Librosa to extract a total of 19 acoustic features (tempo, beats,

chroma stft, chroma cq, chroma cens, melspectrogram, mfcc, mfcc delta, rmse, spectral bandwidth, cent, contrast, rolloff, poly features, tonnetz, zero crossing rate, harmonic, percussive, frame etc.). Among the features provided by Librosa, low-level features like MFCC and ZCR are the most used in the MIR system [50], and mid-level features like chroma represent a meaningful classification of pitch. Not only is this useful for music analysis, but it is also suitable as a feature for effect classification or recognition models [51].

3.3.2. Music Mood Recognition Learning Based on XGBoost Algorithm and Model Production

XGBoost [52] uses a boosting method that utilizes several shallow depth determination tree classifiers. After producing many determination tree classifiers which process the learning method to increase classification accuracy while finding acoustic features which can well classify the moods among acoustic features extracted from classical music segment data, learning is processed in the order of each determination tree. If the result of learning the first decision tree classifier does not classify the mood well, the second decision tree is learned by utilizing a weighted factor. After reflecting the weighted factor according to the test result of the secondary decision-making tree. Then, a third decision making tree is learned by reinforcement. According to the above stipulated process, XGBoost has excellent recognition performance, gradually lowering errors by sequentially reinforcing training data in order. In addition, XGBoost accelerates the learning speed with parallel and distributed computing, so it performs a model search faster to aid a faster processing of large datasets. This study uses XGBoost to create 18 binary classification models for each mood class that is classified as a positive class or a negative class. Since the listener should feel a rich mood, i.e., a complex mood after listening to the music segment, the music segment should be multi-label classified into different class labels of mood. Therefore, in this study, multi-label classification was found to be more useful by using a classification model for each affect rather than recognizing the mood of the whole music.

4. Evaluation

In order to improve the performance of recognizing the detailed mood flow of classical music, in the previous chapter, the music segment extraction method, which is the unit of mood recognition, was applied and a model capable of classifying 18 moods was produced. In this section, from the model we made earlier, the performance evaluation (of the model) to assess the music segmentation method that can improve the model's mood recognition performance and the evaluation to secure the validity of using the XGBoost algorithm are carried out.

The proposed mood classification model in this study is a binary classification model which performs a five-fold cross validation to prevent overfitting for performance evaluation. The performance evaluation results of the model are, validated by the representative performance evaluation indicators of accuracy, precision, recall, F1 score, and ROC AUC. The receiver operating characteristic (ROC) is an index showing a pair of the percentage of samples predicted to be positive among all positive samples and the percentage of samples that were incorrectly predicted to be positive among all negative samples as a curve. Area under the curve (AUC) is an index indicating the area under the ROC curve, and the closer to 1.0, the better the performance. In this study, performance evaluation is conducted based on the ROC AUC score, which is mainly used as an indicator of the performance of the binary classification.

4.1. Performance Evaluation of Segment Method

4.1.1. Experiment 1: Performance Evaluation Method

In order to confirm the difference in performance according to the difference in music segment data extraction method, an experiment was conducted to compare the performance of dataset A and dataset B. Dataset A generated from segments which is a unit of specific mood recognition of classical music in constant intervals of 30 s. Dataset B was generated by adding an algorithm of combining segments of 5 s or under, front and back, which is

considered to be difficult in terms of mood recognition after segmenting music based on MFCC acoustic features analysis.

As the mood class per labelled music segment is not balanced, the data size is selected based on mood class data with the fewest labels so that the experiment can be conducted in similar conditions. The ratio of positive and negative classes of training data for each mood is configured to be a 1:1 ratio to establish a balanced data. The threshold of the model is set to 0.5. If the performance of each model is 0.5 or higher, it is classified as a positive class. In this way, one segment can be classified into several moods such as "Energetic" and "Powerful". Next, to verify the performance of the mood recognition algorithm, the optimal algorithm is selected based on the average of the AUC values among the results classified in each fold using the K-fold cross validation algorithm.

4.1.2. Experiment 1: Result and Discussion

After comparing the Area Under the Curve (AUC) points to assess the performance of Dataset A and Dataset B in Table 2 and Figure 9, Dataset B had higher AUC points at 15 classes (energetic, powerful, joyous, aroused, scary, tense, sad, soft, restrained, happy, mysterious, elegant, tuneful, majestic, reverent) among the 18 mood classes. However, in three mood classes (relaxed, calm, warm), the results of Dataset A were higher. Meanwhile, joyous had the highest performance of 0.89 in Dataset B. Scary had the lowest performance of 0.41 in Dataset A.

Table 2. Dataset A, Dataset B Model Experiment Result.

Mood	Accuracy		Precision		Recall		F1		ROC-AUC	
	A	B	A	B	A	B	A	B	A	B
Energetic	0.68	0.74	0.69	40.72	0.71	0.78	0.69	0.74	0.72	0.77
Powerful	0.45	0.78	0.50	0.80	0.41	0.77	0.42	0.78	0.48	0.85
Joyous	0.58	0.84	0.58	0.85	0.64	0.87	0.60	0.85	0.56	0.89
Aroused	0.56	0.66	0.55	0.67	0.59	0.68	0.55	0.67	0.63	0.68
Scary	0.47	0.65	0.49	0.65	0.57	0.64	0.53	0.64	0.41	0.70
Tense	0.43	0.62	0.43	0.61	0.43	0.68	0.42	0.64	0.49	0.56
Relaxed	0.59	0.56	0.59	0.57	0.61	0.50	0.60	0.53	0.65	0.55
Sad	0.53	0.50	0.53	0.49	0.54	0.48	0.52	0.48	0.54	0.56
Soft/Quiet	0.57	0.60	0.59	0.59	0.59	0.62	0.57	0.60	0.66	0.69
Restrained	0.68	0.63	0.74	0.61	0.59	0.70	0.65	0.64	0.66	0.67
Calm	0.60	0.60	0.61	0.60	0.61	0.54	0.60	0.56	0.57	0.62
Happy	0.55	0.61	0.55	0.62	0.59	0.59	0.56	0.59	0.55	0.70
Warm	0.60	0.54	0.60	0.54	0.64	0.62	0.61	0.57	0.61	0.56
Mysterious	0.50	0.53	0.49	0.54	0.48	0.49	0.47	0.49	0.52	0.53
Elegant	0.50	0.57	0.49	0.57	0.53	0.57	0.49	0.57	0.48	0.55
Tuneful	0.58	0.66	0.58	0.63	0.61	0.65	0.59	0.63	0.63	0.66
Majestic	0.62	0.72	0.61	0.74	0.68	0.71	0.64	0.72	0.67	0.73
Reverent	0.52	0.75	0.52	0.73	0.57	0.80	0.54	0.76	0.58	0.74

There were no mood classes over 0.8 in Dataset A but there were mood classes including Powerful and Joyous over 0.8 in Dataset B, so the mood classification model of Dataset B showed better performance. Based on these results, Dataset B which analyzed moods via the segment extraction method showed a better performance compared to Dataset A which was extracted randomly by constant interval. Moreover, powerful, scary, and tense of Dataset A did not prove higher than 0.5. If AUC is not higher than 0.5, the most likely explanation was that the labelling is wrong. After checking the labelling data, there were many moods corresponding to 30 s, so the performance of the model is predicted to be low. Based on these results, it was shown that Dataset A is unsuitable for the purposes of model learning.

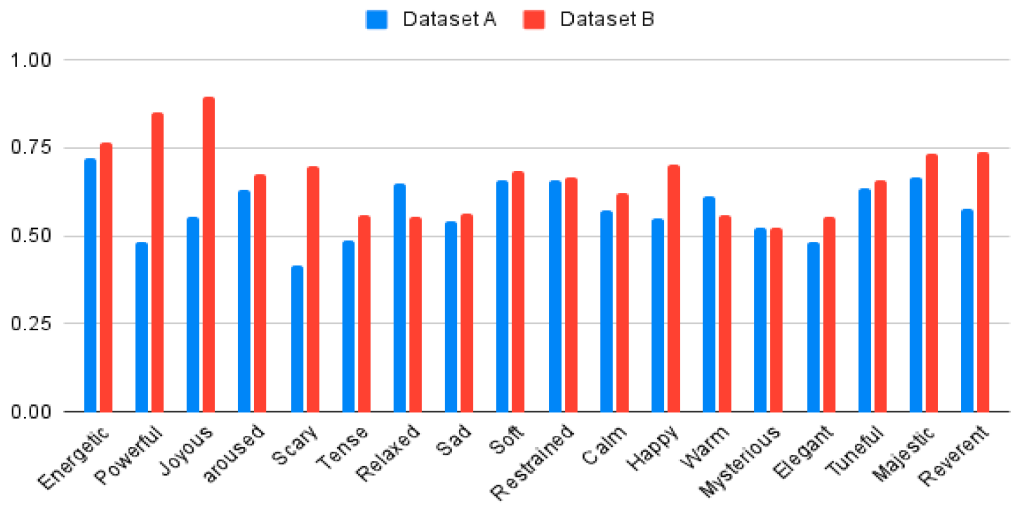

Figure 9. Comparison of ROC-AUC Performance Results of Dataset A and Dataset B.

4.2. Algorithm Performance Evaluation of Proposed Model

4.2.1. Experiment 2: Performance Evaluation Method

The research performs learning of data labelled by music experts using not only XGBoost, which was used for classical music mood recognition model produced by the research, but also random forest and SVM algorithms. Using the validation data, the objective is to derive the performance values between algorithms so that a comparative analysis can be executed. The objective of this experiment is to compare the use of the XGBoost algorithm used for the music mood recognition model with other algorithms, and to establish the foundation for the use value of the XGBoost algorithm in the music mood recognition model. First, 18 binary classification models are produced, classifying whether 18 mood adjectives belong to a positive class or negative class per algorithm, i.e., XGBoost, random forest, and SVM. In general, because the listener experiences a plethora of moods, that is, complex moods throughout the music segment, the music segment should undergo multi-label classification so that it can be classified into several mood class labels. To accomplish this, a model is used for each mood class.

Meanwhile, the ratio of positive and negative classes of training data for each mood is configured by 50% to establish a balanced data. Then, the threshold of the model is set to 0.5, and if the performance of each model is 0.5 or higher, it is classified as a positive class. In this way, one segment can be classified into several moods such as "Energetic" and "Powerful". Next, in order to verify the performance of the mood recognition algorithm, we propose an optimal algorithm based on the average of the AUC values among the results classified in each fold using the K-fold cross validation algorithm.

4.2.2. Experiment 2: Result and Discussion

Table 3 is a table comparing the performance of each of the three algorithms, and a visual comparison of the results is shown in Figure 10. In 14 moods (energetic, powerful, joyous, aroused, scary, tense, soft, restrained, calm, warm, mysterious, elegant, tuneful, majestic), excluding relaxed, happy, sad, and reverent, the performance of XGBoost was higher than SVM and random forest with a slight difference. Particularly, the powerful mood in XGBoost was 0.822911, showing the highest performance. However, lowest performance was scary from SVM by 0.635494.

Table 3. Cross Validation Result Comparison of SVM, Random Forest, and XGBoost Algorithm.

	SVM					Randomforest					XGBoost				
Mood	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC-AUC**	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC-AUC**	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC-AUC**
Energetic	0.71	0.69	0.76	0.72	0.78	0.72	0.71	0.73	0.72	0.79	0.72	0.71	0.75	0.73	0.80
Powerful	0.73	0.73	0.73	0.73	0.81	0.74	0.75	0.74	0.74	0.81	0.74	0.75	0.74	0.74	0.82
Joyous	0.70	0.68	0.77	0.72	0.77	0.72	0.71	0.74	0.72	0.78	0.72	0.71	0.76	0.73	0.79
Aroused	0.68	0.67	0.72	0.69	0.75	0.68	0.67	0.70	0.68	0.75	0.68	0.68	0.69	0.68	0.76
Scary	0.55	0.57	0.50	0.53	0.64	0.63	0.67	0.58	0.59	0.65	0.64	0.63	0.66	0.64	0.70
Tense	0.62	0.66	0.52	0.58	0.66	0.64	0.65	0.59	0.63	0.69	0.64	0.65	0.60	0.62	0.69
Relaxed	0.67	0.65	0.75	0.69	0.72	0.68	0.68	0.70	0.69	0.75	0.68	0.67	0.69	0.68	0.75
Sad	0.63	0.66	0.57	0.61	0.67	0.65	0.68	0.64	0.65	0.73	0.66	0.67	0.63	0.65	0.73
Soft	0.72	0.71	0.74	0.73	0.77	0.72	0.71	0.73	0.73	0.79	0.73	0.72	0.75	0.73	0.80
Restrained	0.70	0.69	0.74	0.72	0.77	0.72	0.71	0.73	0.71	0.78	0.71	0.71	0.73	0.72	0.78
Calm	0.63	0.61	0.73	0.67	0.68	0.65	0.65	0.68	0.66	0.70	0.65	0.64	0.69	0.66	0.70
Happy	0.59	0.59	0.60	0.60	0.64	0.65	0.66	0.62	0.64	0.71	0.65	0.65	0.66	0.65	0.71
Warm	0.62	0.62	0.62	0.62	0.68	0.67	0.67	0.63	0.65	0.72	0.66	0.66	0.66	0.66	0.73
Mysterious	0.61	0.63	0.57	0.60	0.66	0.61	0.63	0.60	0.61	0.66	0.63	0.64	0.59	0.61	0.68
Elegant	0.64	0.65	0.63	0.64	0.69	0.66	0.66	0.63	0.65	0.71	0.66	0.66	0.64	0.65	0.71
Tuneful	0.64	0.63	0.69	0.66	0.69	0.65	0.65	0.64	0.66	0.72	0.66	0.66	0.68	0.67	0.73
Majestic	0.70	0.73	0.66	0.69	0.77	0.69	0.69	0.67	0.69	0.75	0.71	0.71	0.70	0.70	0.77
Reverent	0.64	0.67	0.54	0.60	0.70	0.66	0.67	0.62	0.66	0.74	0.69	0.71	0.64	0.67	0.73

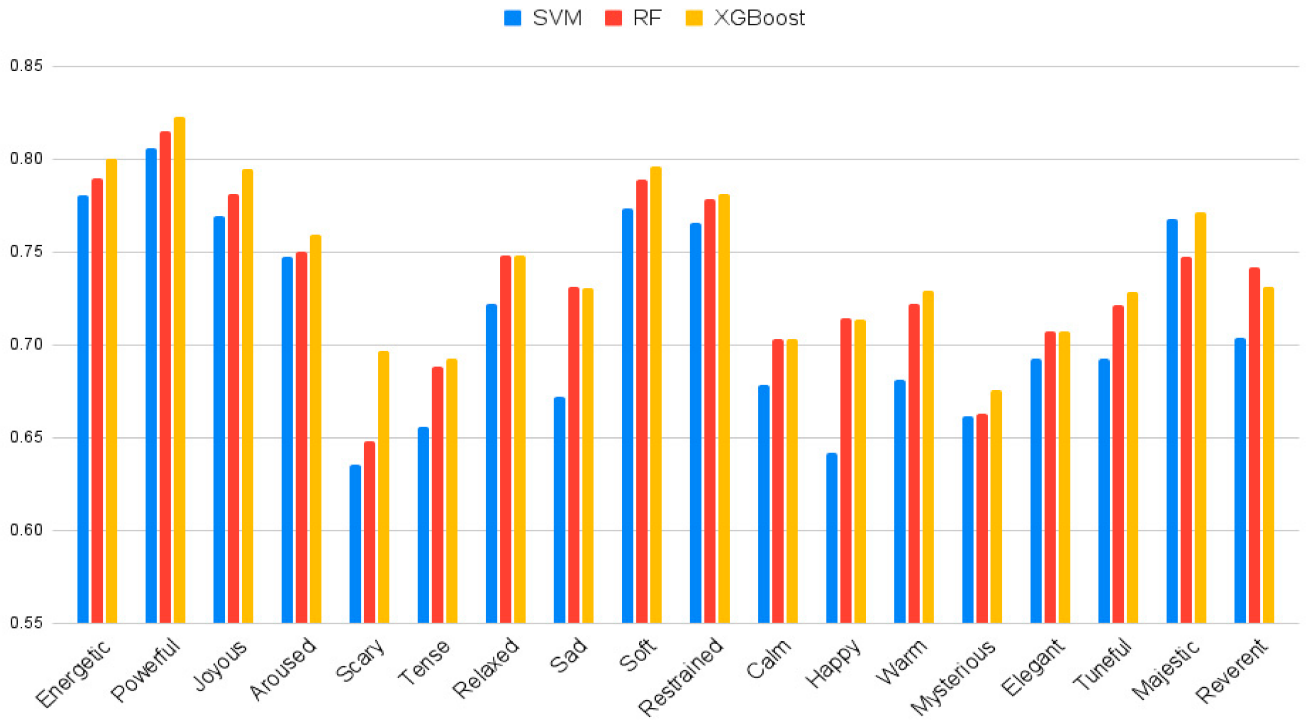

Figure 10. Comparison of ROC-AUC Performance Results Graph of Dataset B by SVM, Random Forest, and XGBoost Model.

5. Conclusions

In this study, a model that automatically recognizes the mood of classical music for a project to create a platform and application service that allows classical music to be fused with visual media suitable for emotions was examined. Classical music was considered to be a good research subject due its inherent characteristic of a long–playing time and transitional flows of various moods. Considering the affect convergence service, the music segment data extraction criteria were established to effectively recognize the various moods detailed in accordance with the flow of the classical music playing time. After generating a model that can automatically recognize the mood according to the acoustic features analysis of each segment data, the performance was verified.

First, more than 12,000 pieces of classical music data that were allowed to be processed were collected. In order to recognize a detailed mood of music data, a standard for dividing the segment data, that is, a length unit of music for recognizing a detailed mood, is first

required. In this study, in order to find a way to divide the music segment data that can improve the performance of mood recognition, the collected music data were segmented in two ways, namely in Dataset A and Dataset B. Dataset A extracts music segment data at regular intervals of 30 s, and Dataset B extracts a segment clustered among acoustic features with similar MFCCs related to tone. The problem with a short music segment of 5 s or under is that it may be a meaningless section for mood recognition, such as the sound of applause, and the time is short in a service that features fusion with visual media, so it is unsuitable for watching visual media. In order to improve this problem, segments of 5 s or under are clustered once again with the front or rear segments, so that all music segments can have a length of at least 5 s. In addition, we derived 18 classes of optimal mood adjectives to be used in the study of the mood recognition model of classical music. Music experts labeled some of music segment data classified as training data for datasets A and B using 18 mood adjectives. Using this as training data, acoustic features were analyzed using XGBoost and a model for recognizing the mood was trained with it. Finally, by using XGBoost to classify mood classes for each music segment, a model was created that automatically detects mood according to the flow of the entire classical music.

Experiments were performed for two evaluations of the classical music emotional model production conducted in this study. A performance evaluation experiment was first conducted to determine the music segment method that can improve the music mood recognition performance by comparing the results from Dataset A and Dataset B. The results revealed that Dataset B showed better recognition results from 15 mood classes (energetic, powerful, joyous, aroused, scary, tense, sad, soft, restrained, happy, mysterious, elegant, tuneful, majestic, reverent) than Dataset A. Through this, it was found that the method from dataset B was optimal for increasing the clarity of mood recognition as a method of clustering and extracting segments of classical music, which is the unit for recognizing the mood in this study. The next performance evaluation experiment involved securing the validity of using the XGBoost algorithm in the model based on the theoretical background in this study. For this purpose, we compared the performance of mood recognition with other SVM and random forest algorithms. The results revealed that XGBoost classified 14 (energetic, powerful, joyous, aroused, scary, tense, soft, restrained, calm, warm, mysterious, elegant, tuneful, majestic) mood classes better than the other two algorithms. Moreover, the ROC AUC values of most mood classes from the model classification results were between 0.7 and 0.8 so that the corresponding classification model is determined to be a good binary classifier. According to the results of model classification, most mood classes have ROC AUC values between 0.7 and 0.8, so it can be considered that the classification model is a good binary classifier.

This study proposed a model that can effectively recognize the detailed mood of classical music, focusing on classical music that has a long playback length and composed of various detailed moods. This model has four important values. First, classical music is segmented according to the tone of 'acoustic features', but it can be segmented with a length optimized for mood recognition. Such a method is a music segment method suitable for not only improving the performance of mood recognition, but also for applications to a service that combines visual appreciation. Second, for the mood classification of classical music, 18 representative mood adjectives were derived from among numerous mood adjectives so that they could be utilized in other similar studies and emotional fusion services. Third, for music segment data to be used as training data, music experts participated in mood labeling to enhance the expertise of training data and results. Finally, after learning the mood of music segment data through the acoustic features of music using XGBoost, a model for recognizing various moods of music segment data was presented to improve mood recognition performance. The model proposed in this study was examined in consideration of the development of an affect convergence platform service based on classical music in the future. In other words, this study can serve as a foundation for efficiently segmenting music and recognizing the mood for a service that allows users to enjoy visual media of the same affect as well as the auditory experience while listening to classical music. In addition,

these researchers intend to expand upon the study of developing a model for recognizing detailed moods specialized in classical music to a model study for recognizing moods specialized for various genres, such as new age, blues, ballads, and jazz in the future.

Author Contributions: Conceptualization, H.K., H.J. and S.L.; methodology, H.K., H.J. and S.L.; software, H.J.; validation, H.K., S.L. and H.J.; formal analysis, H.J.; investigation, H.K., S.L. and H.J.; resources, H.K.; data curation, S.L. and H.J.; writing—original draft preparation, S.L. and H.J.; writing—review and editing, H.K.; visualization, H.K., S.L. and H.J.; supervision, H.K.; project administration, S.L.; funding acquisition, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (MOLIT) and Ministry of Science and ICT(MSIT) (Grant 21NSPS-B159148-04). Also, this work was supported by a research grant from Seoul Women's University (2021-0146).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moore, K.S. A Systematic Review on the Neural Effects of Music on Emotion Regulation: Implications for Music Therapy Practice. *J. Music. Ther.* **2013**, *50*, 198–242. [CrossRef] [PubMed]
2. Lee, Y.S.; Ahn, H.J. The effect of early childhood music drama program using classical music on children's musical creativity. *J. Early Child. Educ.* **2020**, *40*, 39–62. [CrossRef]
3. Castillo-Pérez, S.; Gómez-Pérez, V.; Velasco, M.C.; Pérez-Campos, E.; Mayoral, M.-A. Effects of Music Therapy on Depression Compared with Psychotherapy. *Arts Psychother.* **2010**, *37*, 387–390. [CrossRef]
4. McCraty, R.; Barrios-Choplin, B.; Atkinson, M.; Tomasino, D. The Effects of Different Types of Music on Mood, Tension, and Mental Clarity. *Altern. Ther. Health Med.* **1998**, *4*, 75–84.
5. Chi, J. Influence of classical music on the psychological state of college students under stress. *Rev. Argent. De Clínica Psicológica* **2020**, *29*, 906. [CrossRef]
6. Lu, L.; Liu, D.; Zhang, H. Automatic Mood Detection and Tracking of Music Audio Signals. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 5–18. [CrossRef]
7. Tzanetakis, G.; Cook, P. MARSYAS: A Framework for Audio Analysis. *Org. Sound* **2000**, *4*, 169–175. [CrossRef]
8. Cabrera, D.; Ferguson, S.; Rizwi, F.; Schubert, E. PsySound3: A Program for the Analysis of Sound Recordings. *J. Acoust. Soc. Am.* **2008**, *123*, 3247. [CrossRef]
9. McKay, C.; Fujinaga, I. jMIR: Tools for automatic music classification. In Proceedings of the 2009 International Computer Music Conference, ICMC 2009, Montreal, QC, Canada, 16–21 August 2009.
10. McFee, B.; Raffel, C.; Liang, D.; Ellis, D.; McVicar, M.; Battenberg, E.; Nieto, O. Librosa: Audio and music signal analysis in Python. In Proceedings of the 14th Python in Science Conference (SCIPY 2015), Austin, TX, USA, 6–12 July 2015; pp. 18–24.
11. Rachman, F.; Sarno, R.; Fatichah, C. Music Emotion Classification Based on Lyrics-Audio Using Corpus Based Emotion. *Int. J. Electr. Comput. Eng. (IJECE)* **2018**, *8*, 1720. [CrossRef]
12. An, Y.; Sun, S.; Wang, S. Naive Bayes Classifiers for Music Emotion Classification Based on Lyrics. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; pp. 635–638. [CrossRef]
13. Tao, L.; Ogihara, M. Toward Intelligent Music Information Retrieval. *IEEE Trans. Multimed.* **2006**, *8*, 564–574. [CrossRef]
14. Han, B.; Rho, S.; Jun, S.; Hwang, E. Music Emotion Classification and Context-Based Music Recommendation. *Multimed. Tools Appl.* **2010**, *47*, 433–460. [CrossRef]
15. Lee, J.-I.; Yeo, D.; Kim, B.-M. Detection of Music Mood for Context-aware Music Recommendation. *KIPS Trans. Part B* **2010**, *17B*, 263–274. [CrossRef]
16. Seo, Y.-S.; Huh, J.-H. Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications. *Electronics* **2019**, *8*, 164. [CrossRef]
17. Wu, B.; Zhong, E.; Horner, A.; Yang, Q. Music Emotion Recognition by Multi-Label Multi-Layer Multi-Instance Multi-View Learning. In Proceedings of the MM 2014—2014 ACM Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 117–126. [CrossRef]
18. Xiao, Z.; Dellandrea, E.; Dou, W.; Chen, L. What Is the Best Segment Duration for Music Mood Analysis? In Proceedings of the 2008 International Workshop on Content-Based Multimedia Indexing, London, UK, 18–20 June 2008; pp. 17–24. [CrossRef]
19. Li, T.; Ogihara, M. Detecting Emotion in Music. In Proceedings of the 4th International Conference on Music Information Retrieval, Baltimore, MD, USA, 26–30 October 2003. [CrossRef]

20. Tavares, J.C.C.; da Costa, Y.M.G. Music Mood Classification Using Visual and Acoustic Features. In Proceedings of the 2017 XLIII Latin American Computer Conference (CLEI), Cordoba, Argentina, 4–8 September 2017; pp. 1–10. [CrossRef]
21. Rong, J.; Li, G.; Chen, Y.-P.P. Acoustic Feature Selection for Automatic Emotion Recognition from Speech. *Inf. Process. Manag.* **2009**, *45*, 315–328. [CrossRef]
22. James, W. II.—What Is an Emotion? *Mind* **1884**, *9*, 188–205. [CrossRef]
23. Ekman, P. An Argument for Basic Emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [CrossRef]
24. Chen, L.; Mao, X.; Xue, Y.; Cheng, L.L. Speech Emotion Recognition: Features and Classification Models. *Digit. Signal Process.* **2012**, *22*, 1154–1160. [CrossRef]
25. Xiaohua, W.; Muzi, P.; Lijuan, P.; Min, H.; Chunhua, J.; Fuji, R. Two-Level Attention with Two-Stage Multi-Task Learning for Facial Emotion Recognition. *J. Vis. Commun. Image Represent.* **2019**, *62*, 217–225. [CrossRef]
26. Giannopoulos, P.; Perikos, I.; Hatzilygeroudis, I. Deep Learning Approaches for facial emotion recognition: A case study on FER-2013. In *Advances in Hybridization of Intelligent Methods*; Springer: Cham, Switzerland, 2018.
27. De, A.; Saha, A. A Comparative Study on Different Approaches of Real Time Human Emotion Recognition Based on Facial Expression Detection. In Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 19–20 March 2015; pp. 483–487.
28. De, A.; Saha, A.; Pal, M.C. A Human Facial Expression Recognition Model Based on Eigen Face Approach. *Procedia Comput. Sci.* **2015**, *45*, 282–289. [CrossRef]
29. Young, A.W.; Rowland, D.; Calder, A.J.; Etcoff, N.L.; Seth, A.; Perrett, D.I. Facial Expression Megamix: Tests of Dimensional and Category Accounts of Emotion Recognition. *Cognition* **1997**, *63*, 271–313. [CrossRef]
30. Donato, G.; Bartlett, M.S.; Hager, J.C.; Ekman, P.; Sejnowski, T.J. Classifying Facial Actions. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 974–989. [CrossRef]
31. Agrawal, S.; Khatri, P. Facial Expression Detection Techniques: Based on Viola and Jones Algorithm and Principal Component Analysis. In Proceedings of the 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, India, 21–22 February 2015; pp. 108–112.
32. Russell, J.A. A Circumplex Model of Affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161–1178. [CrossRef]
33. Yang, Y.-H.; Lin, Y.-C.; Su, Y.-F.; Chen, H.H. A Regression Approach to Music Emotion Recognition. *IEEE Trans. Audio Speech Lang. Process.* **2008**, *16*, 448–457. [CrossRef]
34. Thayer, R.E. *The Biopsychology of Mood and Arousal*; Oxford University Press: New York, NY, USA, 1989.
35. Yang, Y.; Su, Y.-F.; Lin, Y.-C.; Chen, H. Music Emotion Recognition: The Role of Individuality. In Proceedings of the ACM International Multimedia Conference and Exhibition, Augsburg, Germany, 28 September 2007; pp. 13–22. [CrossRef]
36. AllMusic Mood Tag. Available online: https://www.allmusic.com/moods (accessed on 2 September 2021).
37. Tzanetakis, G.; Cook, P. Musical Genre Classification of Audio Signals. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 293–302. [CrossRef]
38. Weihs, C.; Ligges, U.; Mörchen, F.; Müllensiefen, D. Classification in Music Research. *Adv. Data Anal. Classif.* **2007**, *1*, 255–291. [CrossRef]
39. Scaringella, N.; Zoia, G.; Mlynek, D. Automatic Genre Classification of Music Content: A Survey. *IEEE Signal Process. Mag.* **2006**, *23*, 133–141. [CrossRef]
40. Fu, Z.; Lu, G.; Ting, K.M.; Zhang, D. A Survey of Audio-Based Music Classification and Annotation. *IEEE Trans. Multimed.* **2011**, *13*, 303–319. [CrossRef]
41. Bayu, Q.D.P.; Suyanto, S.; Arifianto, A. Hierarchical SVM-KNN to Classify Music Emotion. In Proceedings of the 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 5–6 December 2019; pp. 5–10. [CrossRef]
42. Babu, P.A.; Siva Nagaraju, V.; Vallabhuni, R.R. Speech Emotion Recognition System With Librosa. In Proceedings of the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 18 June 2021; pp. 421–424.
43. FMA: A Dataset for Music Analysis. Available online: https://github.com/mdeff/fma (accessed on 2 September 2021).
44. Musopen. Available online: https://musopen.org/music/ (accessed on 2 September 2021).
45. MusicNet. Available online: https://homes.cs.washington.edu/~{}thickstn/musicnet.html (accessed on 2 September 2021).
46. KkachilhanClassic. Available online: http://www.kkacl.com/ (accessed on 2 September 2021).
47. URMP. Available online: http://www2.ece.rochester.edu/projects/air/projects/URMP.html (accessed on 2 September 2021).
48. Collins English Thesaurus. Available online: https://www.collinsdictionary.com/dictionary/english-thesaurus (accessed on 25 September 2021).
49. Oxford Learner's Dictionaries. Available online: https://www.oxfordlearnersdictionaries.com/ (accessed on 25 September 2021).
50. Senac, C.; Pellegrini, T.; Mouret, F.; Pinquier, J. Music Feature Maps with Convolutional Neural Networks for Music Genre Classification. In Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, Florence, Italy, 19 June 2017; pp. 1–5.

51. Imran, D.; Wadiwala, H.; Tahir, M.A.; Rafi, M. Semantic Feature Extraction Using Feed-Forward Neural Network for Music Genre Classification. *Asian J. Eng. Sci. Technol.* **2017**, *6*, 1.
52. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]