

# Korean Traditional Music Genre Classification Using Sample and MIDI Phrases

JongSeol Lee<sup>1,2</sup>, MyeongChun Lee<sup>2</sup>, Dalwon Jang<sup>2</sup>, and Kyoungro Yoon<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering  
Konkuk University, Seoul, Korea

[e-mail: {leejs98, yoonk}@konkuk.ac.kr]

<sup>2</sup>Smart Media R&D

Korea Electronics Technology Institute, Gyeong-GI, Korea

[e-mail: {lmcspring, dalwon}@keti.re.kr]

\*Corresponding author: Kyoungro Yoon

*Received September 27, 2017; revised February 14, 2017; accepted March 8, 2017;  
published April 30, 2018*

---

## Abstract

This paper proposes a MIDI- and audio-based music genre classification method for Korean traditional music. There are many traditional instruments in Korea, and most of the traditional songs played using the instruments have similar patterns and rhythms. Although music information processing such as music genre classification and audio melody extraction have been studied, most studies have focused on pop, jazz, rock, and other universal genres. There are few studies on Korean traditional music because of the lack of datasets. This paper analyzes raw audio and MIDI phrases in Korean traditional music, performed using Korean traditional musical instruments. The classified samples and MIDI, based on our classification system, will be used to construct a database or to implement our Kontakt-based instrument library. Thus, we can construct a management system for a Korean traditional music library using this classification system. Appropriate feature sets for raw audio and MIDI phrases are proposed and the classification results—based on machine learning algorithms such as support vector machine, multi-layer perception, decision tree, and random forest—are outlined in this paper.

---

**Keywords:** Music genre classification, music sample analysis, MIDI analysis, Korean traditional music, music editor

---

A preliminary version of this paper was presented at APIC-IST 2017, and was selected as an outstanding paper. This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2017.

## 1. Introduction

Following the spread of digital music, music information retrieval (MIR) research began to receive increasing attention [1]. In this field of research, many challenging tasks exist, including query-by-singing/humming [2, 3], tempo estimation [4], cover-song identification [5], audio fingerprinting [6, 7], audio tagging [8], automatic melody extraction [9, 10], and music genre classification [11-17]. Among them, music genre classification is considered to be one possible way of managing a large digital music database. The musical genre, a high-level descriptor of music, is extensively used in music stores, radio stations, and on the Internet. It gives a general guideline for retrieving music or for separating and organizing music files. Manual classification of music by genre is laborious and time-consuming work [13]. Thus, an automatic music genre classification system is a way to determine the proper genre for music at a low cost.

To accomplish this task, most studies extract features from an audio dataset, or a collection of audio information organized into a group. MIDI is also used as a raw dataset because it is easy to modify and manipulate [16]. MIDI is a symbolic representation method for musical information. Raw audio data represents actual sound as a digital signal, while MIDI represents a musical event as one or two data bytes, according to the MIDI format specification type. The original purpose of MIDI is to communicate with individual instruments and to enable one instrument, like a computer MIDI console, to control another, such as a MIDI-enabled keyboard. A standard MIDI file format, developed by the MIDI Manufacturers Association, consists of a sequence of MIDI messages [18]. Channel messages such as “Note On,” “Note Off,” “Control Change,” etc., are used as the raw features of music genre classification. The biggest difference between audio- and MIDI-based music genre classifications is the feature set type. Timbre, rhythm, and pitch are generally used in audio data. The timbre feature represents the tone quality of instruments, including human voices. One well-known method is Mel-Frequency Cepstral Coefficients (MFCC), extracted by taking the log-amplitude of a magnitude spectrum and then smoothing the grouped fast Fourier transform (FFT) bins according to the perceptually motivated Mel-frequency scaling [19]. In addition, other features that represent speech and audio, such as spectral centroid, zero crossing, and energy, have been suggested in previous studies [20-22]. MIDI features can also represent audio data with some limitation. MIDI cannot represent the detail found in timbre. This is because audio features are extracted from digitalized raw audio, while MIDI features are extracted from command information for instruments, which does not contain detailed musical information. However, MIDI features have a number of advantages when compared with audio features. MIDI is easy to share, store, and edit. The size of the data in terms of computer space is small. Because of its simplicity, computing when extracting the features and analysis of a feature set can be done speedily. MIDI also has a musicological meaning and can be used for more applied applications. Furthermore, one study reported a low correlation between timbre and genre, which means that when determining genre, information like timbre could, in some cases, be worthless [23].

In this paper, we propose a genre classification system for Korean traditional music using MIDI or .wav files. Each file contains a phrase of music within a genre. The phrase is a minimal chunk communicating a musical pattern that has a musical sense, one that is difficult to classify. Two different forms of input (MIDI and wav file) are implemented, and their performances are investigated. To develop the system, we created a Korean traditional music

genre dataset. Because MIR has lacked proper datasets to explore it, Korean traditional music has not yet been used in MIR research. For a music genre classification system to be realized, a proper dataset is essential. Various datasets such as GTZAN and MSD have been released, and among datasets, the GTZAN music genre dataset is the most famous and widely used [24, 25]. Our dataset, which we will release upon completion, is designed for Korean traditional music. Our work is a part of larger project: making virtual instruments for Korean traditional music and web-based music production/playing/distribution platforms in which the virtual instruments can be used. The music clips in the dataset, which were created and labeled by Korean traditional musicians, were originally collected using a small group of virtual instruments. Because the music clips are short, classification is difficult. Our Korean traditional music genre classification system and dataset will be implemented in an authoring tool in the platform, which makes it easier to produce Korean traditional music. This paper is unique because it is the first involving MIR research for Korean traditional music. The genre classification system proposed in this paper showed satisfactory results for our dataset, and these results can be applied as a guideline for Korean traditional music genre research.

This paper is organized as follows. In Section 2, we review previous work in genre classification. Section 3 introduces Korean traditional music and instruments. Section 4 describes the classification method, including the feature extraction method. Section 5 explains and analyses the experimental results of our classification method. Section 6 introduces the prototype applications that are based on Kontakt and the web environment. The result of our work will feed into the application. Finally, we summarize our study and explain our future work in Section 7.

## 2. Related Work

In general, the MIR system has been used with pop music, but traditional music has begun to garner interest [26-29]. In [26], information related to gathering data on flamenco music from general knowledge bases is shown. As a result of this work, a new knowledge base containing information about flamenco music was created and released. Using this knowledge base, artist relevance can be computed. From this research, we can consider that only gathering and arranging the music information can make a new value with information retrieval technique. In [27], a new Greek Music Dataset (GMD) was proposed. The authors of that paper hold the opinion that Greek traditional music presents a number of unique characteristics. The GMD includes information for 1,400 Greek music tracks, arranged by track, lyrics, symbolic features, manually annotated labels, and so on. In [28], a large machine-readable dataset of Turkish makam musical phrases is presented. In the dataset, musical scores are segmented into phrases by experts. The authors of the paper argue that Turkish makam music is rarely studied since data resources are not available. In [29], a classification for phrase-based Indian art music, rāga, was suggested. They collected 124 hours of audio in 480 recordings, including 40 of Carnatic music, for use in the genres of rāga. They extracted a feature matrix using three steps (melodic pattern discovery, melodic pattern clustering, and feature extraction). Some classification algorithms were compared for evaluation and a naïve Bayes classifier outperformed the other. As written in the papers [26 -29], a database is essential for MIR research, so the researchers created their own databases. The databases are explained in **Table 1**. Through these attempts, we can see that existing MIR techniques, based on collected data, contribute to the development and investigation of traditional culture. There is little in the way of Korean traditional music datasets for MIR research. On the Traditional Korean Music site

[30] and Gugak<sup>1</sup> archive from the National Gugak Center [31], music clips and performance video clips of Korean traditional music are provided, but it is difficult to use these clips in MIR research.

**Table 1.** Traditional music database [26-29]

Music	Data type	Amount
Flamenco[26]	Audio, Artist, Album, Track, Palo and Place, three domain specific classes, and so on	14,078 tracks
Greek[27]	Audio, lyrics, symbolic features, mood, genre, MIDI, and so on	1,400 tracks
Turkish makam music [28]	Symbolic	31,362 phrases
Indian art music [29]	Audio	124 hours of audio in 480 tracks

A music genre classification system has two parts: feature extraction and classification based on the feature set [11]. The input can be a MIDI file or wav file. Most music genre classification systems use wav files because it is more difficult to collect MIDI files. In some applications, MIDI files converted from wav files have been used [32], but to the best of our knowledge, converted MIDI files have not been used in music genre classification. However, whether using MIDI or wav as an input is not the problem of difficulty. MIDI input can lead to the easy extraction of musical features related to both music structure and notes, but it cannot supply information about timbre feature. Some genres have special properties related to timbre. Thus, whether MIDI input is better than wav input cannot be concluded, and investigation about comparison of two inputs is not yet announced.

In musical genre classification system with wav input, various features which are obtained from signal processing techniques are used: spectral feature, MFCC, octave-based spectral contrast (OSC), octave-based modulation spectral contrast (OMSC), decorrelated filter bank (DFB), and so on. The feature values themselves, statistical values (mean, median, maximum, standard deviation, and so on) of the feature values, and variations and/or modulations of the feature values are used as input vectors for the classifiers. A spectral feature is the value calculated from the output of short-time Fourier transform (STFT) such as spectral centroid, spectral roll off, zero crossing, and others [11]. MFCC refers to perceptually motivated features that are also based on the STFT [11]. It has been widely used in various speech processing areas, and it represents spectral characteristics based on Mel-frequency scaling [33]. Many articles have proven the effectiveness of MFCC as a main tool for music classification. OSC involves the spectral peak, spectral valley, and spectral contrast in each octave-based sub band [34]. DFB involves the variation of amplitudes between neighboring bands. It was originally used for speaker recognition, and it is applied in music genre classification [35]. It is extracted using subtraction of the log spectrum of the neighboring Mel-scale band. OMSC is extracted using long-term modulation spectrum analysis to represent the time-varying behavior of music signals [12]. Through the latest advances in neural network research, a novel method based on neural networks is also proposed in [15].

In musical genre classification systems using MIDI input, information including tempo, note, and rhythm, among others, are used as feature sets. Statistical values/histograms/moments of the features were used in previous studies [16, 17]. A similarity measurement of sequences,

---

<sup>1</sup> *Gugak means Korean traditional music.*

SMBGT (Subsequence Matching with Bounded Gaps and Tolerances) has been used with  $k$ -Nearest Neighbor to classify MIDI data sets [36]. In this study, the researchers collected 100 MIDI files in four genres: classical, blues, rock, and pop. In [37], they focused on transformation of the pitch and duration feature. Four methods, transposition transformation, augmentation/diminution transformation, sequential modulation transformation, and crab transformation, are defined to preserve feature invariance. A combination of audio and MIDI features was introduced in [38]. Well-known feature extraction methods such as MFCC, spectral centroid, spectral roll-off, spectral flux, and time domain zero crossing were used for the audio feature. They did not extract MIDI features but measured the distance between two music pieces using normalized compression distance (NCD). The result of this voting method showed an improvement in accuracy.

### 3. The Korean Traditional Music Genre

As shown in Fig. 1 we concentrate on four genres: "Minyo", "Sanjo," "Gagok," and "Pungryu." We also examine 12 subgenres: Jinyangjo, Jungmori, JungJungmori, Jajinmori, Whimori, Gutgeri, Gagok, Pungryu, Gyeonggi, Sedo, Dongbu, Namdo, Gagok, and Pungryu are used for our study

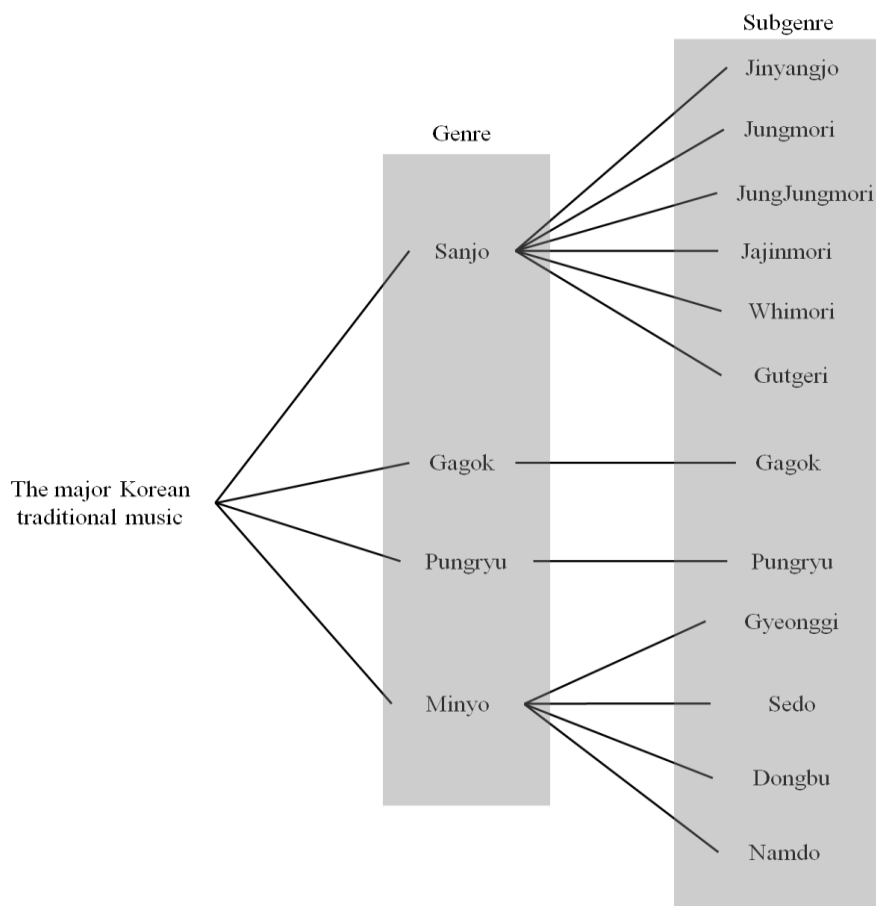


Fig. 1. Taxonomy of the Korean traditional music genres for the proposed system

Minyo is a song passed down orally, bearing a traditional set of rhythms. Minyo has been handed down, separated by natural geography, preserving their unique colors. Currently, the transmission of Minyo is divided into four districts with distinct musical characteristics: Gyeonggi, Seodo, Dongbu, and NamDo. Therefore, the geographical subdivision of Minyo is similar to genre classification in western music.

Sanjo is an instrumental type of music in which the tempo increases for each movement. Sanjo is typical Korean instrumental music and is composed of the plural music movements. Each movement has a subcategory organized by tempo, instrument, or clique. **Table 2** shows the Sanjo subgenres.

**Table 2.** The subgenres of Sanjo

Subgenre	Tempo
Jinyangjo	Slowest
Jungmori	Slower, 12/4 time
JungJungmori	Between Joongmori and Jajinmori
Jajinmori	Faster, 12/8 time
Gutgeri	Faster, 12/8 time
Whimori	Fastest

Sanjo is a branch of traditional music based on Sanna's roots in shaman culture and Pansori. It developed as a folk art and has since been transformed into a high-level instrumental solo performance. Sanjo has a variable characteristic: improvisation, allowing performers to play or to add a melody or a rhythm depending on the whim of the performer. However, even if the improvised music is highly flexible, it is based on logic in its overall structure and form. Gagok is a genre of traditional Korean vocal music sung by men and women accompanied by a small orchestra. Traditional Gagok originated from a song sung among the scholars of the Joseon Dynasty period, and it is distinctly different from recent art songs composed in the western technique. Pungryu is a specific instrument-centric music genre that can be divided into two categories: chordophone-based Pungryu and wind instrument-based Pungryu. It is also called Bangjungak, meaning music for indoor performance. In this paper, we gathered audio samples and MIDI datasets for Korean traditional music. We extracted MIDI datasets from four distinct genres that are heavily utilized in traditional music.

#### 4. Classification Method

Music genre classification is commonly composed of two steps - feature extraction and classification. **Fig. 2** shows the overall architecture of the proposed system.

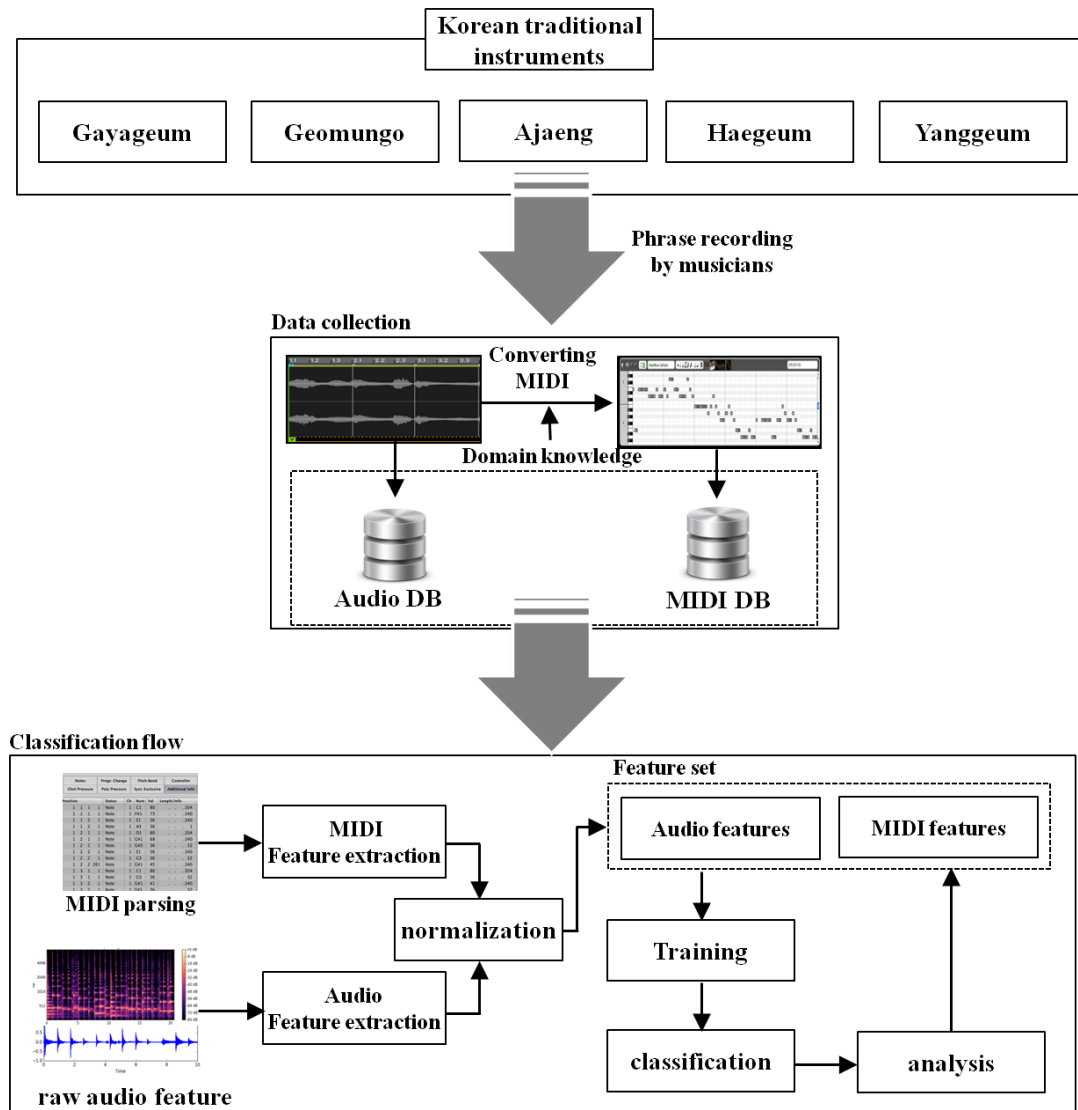


Fig. 2. Overall system architecture

As shown in Fig. 2, each audio sample is recorded by performers using Korean traditional instruments. Gayageum and Geomungo are plucked string instruments. These instruments are played by striking or plucking the strings with a rod-shaped bamboo plectrum. Ajaeng, a zither with silk strings, was designed to perform effect music in the mid-20th century. It has the lowest range of all Korean traditional instruments. It sounds a bit rough and clumsy, but majestic. Haegeum, a fiddle-like instrument, is the instrument that can move about the scale most freely among Korean traditional instruments. Yanggeum is a metal-stringed instrument; it was the only metal instrument played in Korea before the 20th century. It is used for various types of music because of its distinctively clear sound.

The creation of a MIDI data set is done by converting and optimizing the audio samples. To extract the phrases of the entire music, we defined some attributes as shown in Table 3.

**Table 3.** The criteria of phrases gathering

Classification	Sub-classification
Universality	Originality, repetitiveness, conventionality
Functionality	Circularity, correspondence to western music, symbolism, segmentation
Esthetic	Creativity, artistic value

Before we extracted the audio features, we first parsed MIDI events from the MIDI files. A MIDI event includes Note On, Velocity, Control, Change, Tick, etc. It consists of a sequence form. During the feature extraction step, we extracted features as follows:

**Table 4.** MIDI feature set and description

Raw feature vector	Statistical value	Description
Pitch	Mean	The average and standard deviation of MIDI pitch values
	Standard deviation	
	Maximum	Maximum and minimum of MIDI pitch values
	Minimum	
	Rate of change	The average of the absolute value of the difference between Note On values
Velocity	Mean	The average and standard deviation of MIDI velocity values
	Standard deviation	
Tick	Mean	The average and standard deviation of MIDI tick values
	Standard deviation	
Tempo	Initial tempo	Initial tempo and the change in tempo
	Change in tempo	
Control Change	The number of control change	the number of control changed

Audio features were extracted using the method in [Table 5](#) with LibROSA [39]. The features were defined using 12 methods (137 dimensions) and each method was categorized into three types (time related, spectral related, and chroma related). Onset and beat features were estimated from the time intervals of detected onsets and beats, respectively. Spectral features represent the distributions of energy over a set of frequencies. It is composed of 60 dimensions. MFCC represents spectral characteristics, based on Mel-frequency scaling. We extracted 20 MFCC sequences for the raw feature vector. Tonnetz represents tonal centroid features. Six dimensions of features were extracted for each frame. The other well-known features, zero crossing, roll-off, spectral centroid, root mean square energy (RMSE), are used for spectral representation of the audio signal. The chroma feature, also known as chromagram, refers to the twelve pitch classes. We utilized the chromagram of STFT, constant-q (CQT), and chroma energy normalized (CENS).



**Table 5.** Audio feature set and description

Type	Raw feature vector	Statistical value	Dimension
Time related	Tempo	BPM	1
	Onset	Mean and Standard deviation	2
	Beat	Mean and Standard deviation	2
Spectral related	MFCC	Mean and Standard deviation	40
	Tonnetz	Mean and Standard deviation	12
	Zero Crossing	Mean and Standard deviation	2
	Roll-off	Mean and Standard deviation	2
	Spectral Centroid	Mean and Standard deviation	2
	RMSE	Mean and Standard deviation	2
Chroma (Pitch) related	STFT	Mean and Standard deviation	24
	CQT	Mean and Standard deviation	24
	CENS	Mean and Standard deviation	24

The extracted statistical values of raw vectors were normalized, then learned by machine-learning algorithms such as the support vector machine (SVM), multi-layer perception (MLP), random forest (RF), and decision tree (DT). We will compare the performance of the classifiers through the experiments in the next section. SVM, in particular, is well known in genre classification because it has shown good performance with high dimensional features.

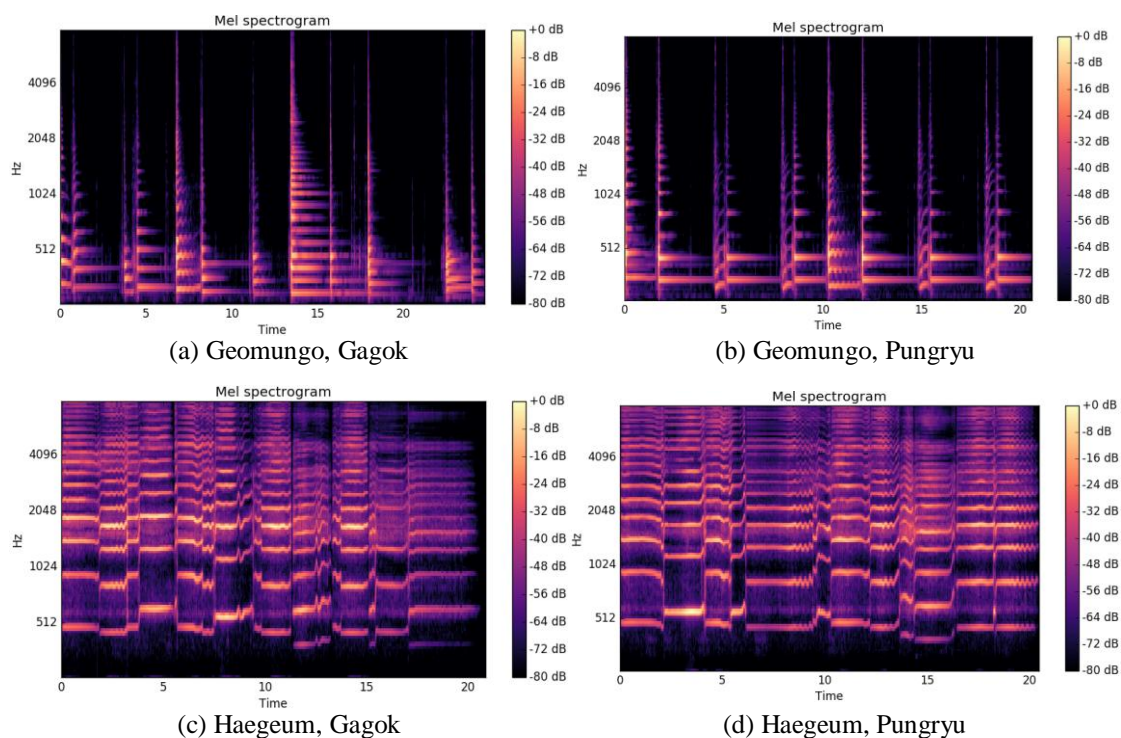
SVM kernels differ in training time, training performance, and overfitting/underfitting problems, depending on the type and size of the data. In our case, we used the linear, RBF, and poly kernels to perform comparative experiments. The size of the hidden node and hidden layer of the MLP was also experimentally determined in a small range. Stochastic gradient descent (SGD) was used as a learning algorithm. Recently, as the performance of deep learning has dramatically improved for large datasets, the number of studies that apply a deep learning algorithm to the genre classification problem have been increasing [40, 41]. However, we did not use deep learning because our data set is small. The basic RF and DT model of Scikit-learn [43] was also used.

## 5. Experimental Results and Analysis

The purpose of the experiment was to determine the feature set and learning algorithm that have the highest performance for four music genres and 12 subgenres. Before the experiment, we performed a simple analysis by visualizing the features of the raw audio data. Next, we compared the performance of the combination of audio features and MIDI features. Finally, a detailed analysis of the accuracy rate was performed through the confusion matrix for the 12 subgenres.

In our experiments, 587 MIDI and audio phrases were used for genre and subgenre classification, respectively. To achieve accuracy in the classification, 10-fold cross validation was used. This means that the model was trained using 9 out of 10 folds as training data, and the result model was validated based on the remaining part of the data. We utilized Scikit-learn for classification and the parameter tuning of all classifiers was done experimentally.

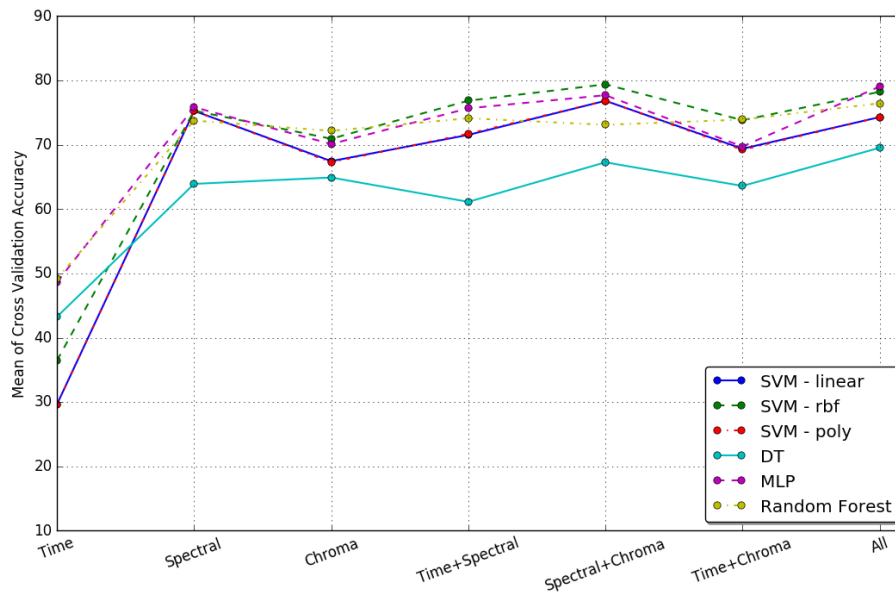
**Fig. 3** is a set of examples of Mel-scaled spectrograms for raw audio phrases. As seen in the figure, the similarities between instruments seem to increase between genres. This may cause performance degradation due to timbre features. We have confirmed this point through the following experiments.



**Fig. 3.** Mel-scaled spectrogram of the audio phrase

First, we measured accuracy based on the feature type defined in **Table 5** to measure the performance difference between the combinations of audio feature types. **Fig. 4** shows the result of classifying four genres by combining audio features. The accuracy of each feature type was the highest for spectral-related features and the performance of all classifiers, except DT, was nearly the same. When combining two feature types, the combination of spectral and

chroma showed the highest accuracy, and SVM-RBF showed the highest performance, with 79.3%. When all features were combined, MLP showed the highest accuracy at 79.0%, but this accuracy was slightly lower than that of the combination of spectral and chroma (SVM-RBF). Considering the relationship between classifiers and feature combinations, DT and MLP showed the highest accuracy combining all features, while SVM and RF are spectral + chroma and tempo + spectral, respectively.



**Fig. 4.** Accuracy of the audio feature combination (4 genres)

The following outlines some important combinations of MIDI features from the preliminary experiments and feature analysis. The accuracy rate of combining MIDI features is shown in [Fig. 5](#), which displays the result of classifying the four genres. The accuracy when using all the features was the highest in all classifiers. The most remarkable aspect in the results is that RF and DT showed the highest accuracy, more than 98% in all feature combinations. They even maintained high accuracy and showed higher performance than other classifiers when single-feature types like velocity or tempo were used. In the results, except for DT and RF, SVM-RBF with all features showed the highest performance, at 98.4%. In terms of feature combinations, tick + tempo and tempo + velocity + pitch combinations were degraded in most classifiers.

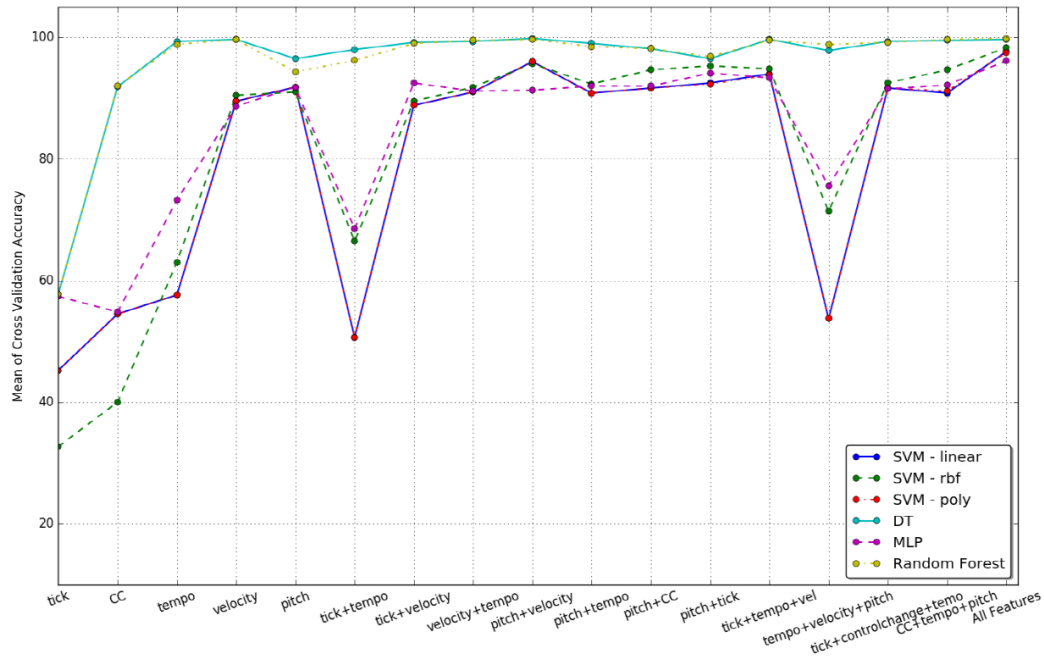


Fig. 5. Accuracy of the MIDI feature combination (4 genres)

Table 6 shows the performance comparison of audio and MIDI features for 12 subgenres. Although the result using the audio feature type showed a significant decrease in accuracy, it maintained high accuracy in some classifiers (SVM-RBF, DT, RF) using MIDI features. It is encouraging that high performance was obtained in some classifiers, even though there was little data to learn from. In addition, most classifiers displayed high performance when all features were used.

Table 6. The average accuracy of 12 subgenres

Type	Classifier	Feature	Accuracy
Audio	SVM(Linear)	All	0.517±0.18
	SVM(RBF)	All	0.544±0.18
	SVM(Poly)	All	0.517±0.18
	MLP	Timbral+Pitch	0.495±0.18
	Decision Tree	Timbral+Tempo	0.376±0.13
	Random Forest	Timbral+Tempo	0.476±0.17
MIDI	SVM(Linear)	All	0.503±0.11
	<b>SVM(RBF)</b>	<b>All</b>	<b>0.883±0.08</b>
	SVM(Poly)	CC+Tempo+Pitch	0.500±0.15
	MLP	All	0.564±0.14
	<b>Decision Tree</b>	<b>All</b>	<b>0.959±0.07</b>
	<b>Random Forest</b>	<b>All</b>	<b>0.969±0.04</b>

**Table 7** is the confusion matrix of MIDI-based subgenre classification; RF was used for the classifier. The performance was measured using the Weka [44] tool to determine the reliability of the previous experimental results. As shown in the table, most subgenres were well classified, at 95.4 and 97.9%, respectively. In particular, the accuracy of RF and DT was 97.9% and 95.4%, respectively, and the performance of MLP was lower than RF and DT.

**Table 7.** Confusion matrix of MDI based subgenre classification

	DB	GG	KG	ND	PR	SD	GI	JJM	JYJ	JJM	JM	WM
DB	<b>28</b>	0	0	0	0	1	0	0	0	0	0	0
GG	0	<b>42</b>	0	0	0	0	1	0	0	0	0	0
KG	0	0	<b>53</b>	0	0	0	0	0	0	0	0	1
ND	0	0	1	<b>38</b>	0	0	0	0	0	0	0	0
PR	0	0	0	0	<b>95</b>	0	0	0	0	0	0	0
SD	0	0	0	0	0	<b>20</b>	0	0	0	0	0	0
GI	0	0	0	0	0	0	<b>28</b>	0	0	0	0	0
JJM	0	0	0	0	0	0	0	<b>70</b>	0	0	1	0
JYJ	0	0	0	1	0	0	0	0	<b>54</b>	0	1	0
JJM	0	0	1	0	0	0	0	0	0	<b>54</b>	0	0
JM	0	0	0	0	0	0	0	0	0	0	<b>56</b>	0
WM	0	0	1	0	0	0	1	0	0	0	0	<b>37</b>

## 6. Prototype Platform and Application

We are using the information in this study to advance our progress in making virtual instruments for Korean traditional music and designing a web-based music production/playing/distribution platform [42]. The proposed Korean traditional music genre classification system will be adopted in the platform with the virtual instruments. Using the genre classification system in the platform led to three benefits. First, our study of a Korean traditional music genre classification system can help people who have an interest in Korean traditional music but who are not familiar with it. In the forthcoming production platform, virtual instruments for Korean traditional music will be created, and every user can use them. However, not every user will be familiar with or understand virtual instruments. Genre-based information and varied examples can help the unfamiliar user. To accomplish this, Korean traditional music genre classification is necessary. Varying music clips related to Korean traditional music are collected in the platform and classified. The clips are intended for the users. Using the results of the classification method and experiment, we have been constructing a sample audio and MIDI database as well. As shown in **Fig. 6**, we are implementing a Kontakt-based Korean traditional instrument library. Kontakt is a virtual studio technology (VST) sampler that can develop a sampler instrument using its own language. The purpose of the prototype application is to construct as large a dataset of Korean traditional music phrases as possible. As the phrase dataset increases, the automatic genre classification process will be increasingly essential.

Second, Korean traditional music lessons can be given in the music platform using the classification system. As written in [42], a lesson service will be opened on the platform. This service was originally planned for piano lessons, but it is not limited to piano. Lessons related to Korean traditional music can be very helpful for building interest in the platform. The

knowledge captured in the classification system and the enlarged database will show the variety of Korean traditional music within the platform. Third, the outputs of the music production platform can be classified appropriately when they are distributed. Music clips produced on the platform shall be sold with proper tags to explain the clips. Traditional music is not familiar to many listeners; thus, correct genre information is essential to give advice on selection.

The features of the prototype application are as follows:

- **Single sounds playing:** A feature that can play the individual sound of Korean traditional instruments, such as Keomungo, Ajang, Gayageum, Haegeum, etc.
- **Phrases playing:** A feature that can load and play constructed phrase sets from Korean traditional music. The phrases in the pattern slots are loaded from categorized datasets from the phrase list.
- **MIDI editor:** A feature that can load, store, and edit constructed MIDI sets for Korean traditional music.

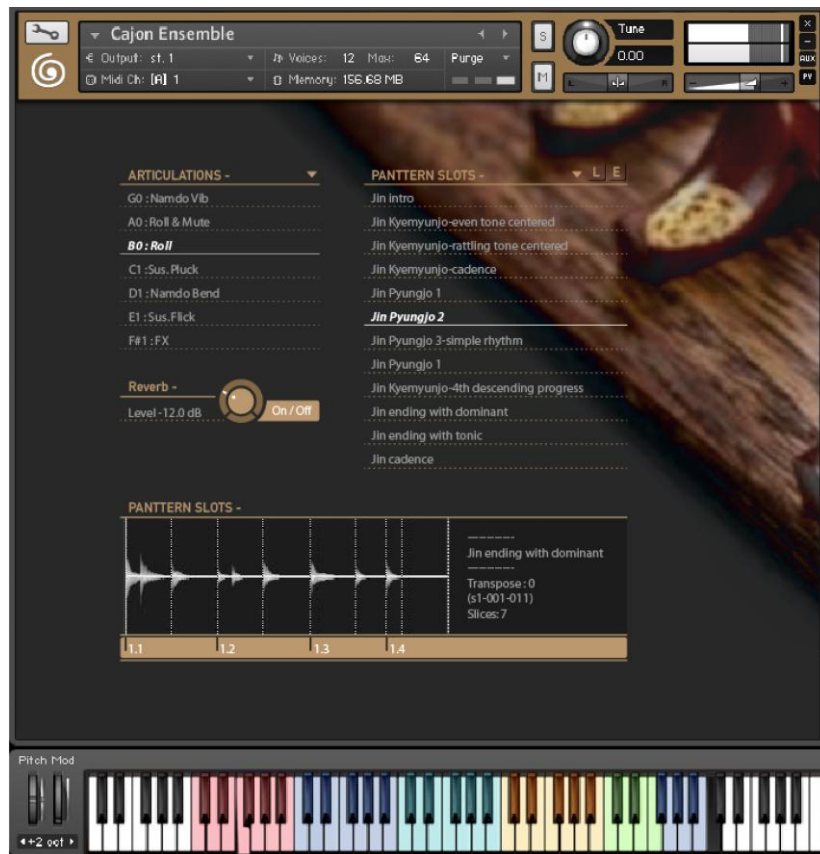


Fig. 6. Prototype application

## 6. Conclusion

In this study, we presented a method to classify phrases of Korean traditional music by genre and subgenre. We classified four genres (Minyo, Sanjo, Gagok, and Pyongyang) and 12 subgenres of Korean traditional music and recorded performances on five traditional musical instruments to collect audio datasets. We then converted them to MIDI format. Datasets were

collected in phrases rather than whole pieces or songs. In the feature extraction step, we extracted 12 features from the MIDI data and 137 features from the audio data. RF showed the highest accuracy compared to other algorithms.

We also introduced the platform and prototype applications, in which the results of this study will be utilized. We explained the necessity of a genre classification module and how it will increase the efficiency of our platform.

In the near future, we are planning to collect more datasets. To improve the accuracy rate, we will explore deep learning algorithms, which have become an increasing issue in recent years. Some studies have already attempted to classify music genres using Convolutional Neural Network (CNN) and have shown promising results. One of the benefits of deep learning is that end-to-end learning is possible, which means that the hand-crafted feature extraction step is not necessary. In [37], CNN outperformed the hand-crafted feature-based method. Furthermore, the combination of CNN with hand-crafted feature extraction showed the highest accuracy rate. Recurrent Neural Network (RNN), one of the deep learning algorithms, can learn sequential data. Audio and MIDI data are suitable for use with this algorithm as sequential data, and we will conduct experiments in this vein to improve the performance of music genre classification.

## References

- [1] J. S. Downie, "Music information retrieval," *Annual Review of Information Science and Technology*, 37:295-340, 2003. [Article \(CrossRef Link\)](#)
- [2] D. Jang, C.-J. Song, S. Shin, S.-J. Park, S.-J. Jang and S.-P. Lee, "Implementation of a matching engine for a practical query-by-singing/humming system," *Int. Symp. on Signal Processing and Information Technology (ISSPIT)*, pp. 258-263, 2011. [Article \(CrossRef Link\)](#)
- [3] J. S. R. Jang and H. R. Lee, "A general framework of progressive filtering and its application to query by singing/humming," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 350-358, 2008. [Article \(CrossRef Link\)](#)
- [4] S. W. Hainsworth and M. D. Macleod, "Particle filtering applied to musical tempo tracking," *EURASIP J. Applied Signal Processing*, vol. 15, pp. 2385-2395, 2004. [Article \(CrossRef Link\)](#)
- [5] D. P. W. Ellis and G. E. Poliner, "Identifying cover songs with chroma features and dynamic programming beat tracking," *Proc. Int. Conf. Acoustic, Speech and Signal Processing*, 2007. [Article \(CrossRef Link\)](#)
- [6] D. Jang, C. D. Yoo, S. Lee, S. Kim and T. Kalker, "Pairwise Boosted Audio Fingerprint," *IEEE Trans. Information Forensics and Security*, vol. 4, no. 4, pp. 995-1004, Dec. 2009. [Article \(CrossRef Link\)](#)
- [7] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," *Proc. International Conf. on Music Information Retrieval (ISMIR)*, 2002. [Article \(CrossRef Link\)](#)
- [8] K. Choi, G. Fazekas, and M. Sandler, "Automatic tagging using deep convolutional neural networks," *Proc. International Conf. on Music Information Retrieval (ISMIR)*, 2016. [Article \(CrossRef Link\)](#)
- [9] S. Jo and C. D. Yoo, "Melody extraction from polyphonic audio based on particle filter," *Proc. International Conf. on Music Information Retrieval (ISMIR)*, pp. 357-362, 2010. [Article \(CrossRef Link\)](#)
- [10] V. Arora and L. Behera, "On-line melody extraction from polyphonic audio using harmonic cluster tracking," *IEEE Trans. on Audio Speech and Language Processing*, vol. 21, no. 3, pp. 520 -530, 2013. [Article \(CrossRef Link\)](#)
- [11] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.* vol. 10, no. 5, pp. 293-302, 2002. [Article \(CrossRef Link\)](#)

- [12] C-H. Lee, J-L. Shih, K-M. Yu, and J-M Su, "Automatic music genre classification using modulation spectral contrast feature," *IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2007. [Article \(CrossRef Link\)](#)
- [13] D. Jang, M. Jin and C. D. Yoo, "Music genre classification using novel features and a weighted voting method," *IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2008. [Article \(CrossRef Link\)](#)
- [14] Y.-F. Huang, S.-M. L., H.-Y. Wu, and Y.-S. Li. "Music genre classification based on local feature selection using a self-adaptive harmony search algorithm," *Data & Knowledge Engineering*, vol. 92 pp. 60–76, 2014. [Article \(CrossRef Link\)](#)
- [15] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Transfer learning for music classification and regression tasks," *Proc. International Conf. on Music Information Retrieval (ISMIR)*, 2017. [Article \(CrossRef Link\)](#)
- [16] C. McKay, "Automatic genre classification of MIDI recordings," *Dissertation, McGill University*, 2004. [Article \(CrossRef Link\)](#)
- [17] J. Valverde-Rebaza, A. Soriano, L. Berton, M. C. F. de Oliveira, and A. Lopes, "Music genre classification using traditional and relational approaches," in *Proceedings of Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 259-264, 2014. [Article \(CrossRef Link\)](#)
- [18] International MIDI Association. *MIDI musical instrument digital interface specification 1.0*. 1983.
- [19] A. Eronen, "Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs," *Seventh Int. Symp. on Signal Processing and Its Applications*, vol. 2, pp. 133-136, 2003. [Article \(CrossRef Link\)](#)
- [20] P. Annesi, R. Basili, R. Gitto, A. Moschitti and R. Petitti, "Audio feature engineering for automatic music genre classification," *Proc. of Int. RIAO Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, pp. 702-711, 2007. [Article \(CrossRef Link\)](#)
- [21] D. Chathuranga and L. Jayaratne, "Automatic music genre classification of audio signals with machine learning approaches," *GSTF Journal on Computing (JoC)*, vol. 3 no. 2 pp. 13-24, 2013. [Article \(CrossRef Link\)](#)
- [22] A. Rosner and B. Kostek, "Automatic music genre classification based on musical instrument track separation," *Journal of Intelligent Information Systems*, pp. 1-22. 2017. [Article \(CrossRef Link\)](#)
- [23] J. J. Aucouturier and F. Pachet, "Representing musical genre: A state of the art," *Journal of New Music Research*, vol. 32, no. 11, pp. 83-93, 2003. [Article \(CrossRef Link\)](#)
- [24] B. L. Sturm, "An analysis of the GTZAN music genre dataset," in *Proc. of the second int. ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pp. 7–12, 2012. [Article \(CrossRef Link\)](#)
- [25] T. Bertin-Mahieux, D. P.W. Ellis, B. Whitman, and P. Lamere. "The Million Song Dataset," in *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, 2011. [Article \(CrossRef Link\)](#)
- [26] S. Oramas, F Gómez, E Gómez, and J. Mora, "FlaBase: Towards the Creation of a Flamenco Music Knowledge Base," in *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, 2015. [Article \(CrossRef Link\)](#)
- [27] D. Makris, I. Karydis, and S. Sioutas, "The Greek music dataset," *Proc. of the 16th Int. Conf. on Engineering Applications of Neural Networks*, 2015. [Article \(CrossRef Link\)](#)
- [28] M.K. Karaosmanoğlu, B. Bozkurt, and A. Holzapfel, "A symbolic dataset of Turkish makam music phrase," in *Fourth International Workshop on Folk Music Analysis (FMA)*, 2014. [Article \(CrossRef Link\)](#)
- [29] S. Gulati, J. Serrà, V. Ishwar, S. Sentürk and X. Serra, "Phrase-based rĀga recognition using vector space modeling," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 66-70, 2016. [Article \(CrossRef Link\)](#)
- [30] Traditional Korean Music, Jeollabuk-do, [online] Available: <http://www.koreamusic.org/langen/main.aspx>.
- [31] Gugak archive, National Gugak Center, [online] Available: <http://archive.gugak.go.kr/ArchivePortal/>



- [32] Y. Lin, X. Chen, and D. Yang, D."Exploration of music emotion recognition based on midi," in *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, 2013. [Article \(CrossRef Link\)](#)
- [33] X. Huang, A. Acero and H.-W. Hon, "Spoken Language Processing," *Prentice Hall PTR*, 2001. [Article \(CrossRef Link\)](#)
- [34] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, "Music type classification by spectral contrast feature," in *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME)*, vol. 1, pp. 113-116, 2002. [Article \(CrossRef Link\)](#)
- [35] S.-C. Lim, J. -S. Lee, S.-J. Jang, S. -P. Lee, and M. Y. Kim, "Music-genre classification system based on spectro-temporal feature and feature selection," *IEEE Trans. on Consumer Electronics*, Vol. 58, No. 4, Nov. 2012. [Article \(CrossRef Link\)](#)
- [36] A. Kotsifakos, E. E. Kotsifakos, P. Papapetrou and V. Athitsos, "Genre classification of symbolic music with SMBGT," in *Proc. of the Int. Conf. on Pervasive Technologies Related to Assistive Environments (PETRA)*, no. 44, 2013. [Article \(CrossRef Link\)](#)
- [37] A. Ruppin and H. Yeshurun, "Midi music genre classification by invariant features," in *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, pp. 397–399, 2006. [Article \(CrossRef Link\)](#)
- [38] Z. Cataltepe, Y. Yaslan, A. Sonmez, "Music genre classification using MIDI and audio features", *EURASIP J. Adv. Signal Process*, vol. 2007, 2007. [Article \(CrossRef Link\)](#)
- [39] B. McFee, C. Raffel, D. Liang, D.P.W. Ellis, M. McVicar, E. Battenberg, and O. Nieto "librosa: Audio and music signal analysis in python." *In Proc of the 14th python in science conference*, pp. 18-25. 2015. [Article \(CrossRef Link\)](#)
- [40] Y. M. G. Costa, L. S. Oliveirab and C. N. Silla Jr.c, "An evaluation of Convolutional Neural Networks for music classification using spectrograms," *Applied Soft Computing*, vol. 52, pp. 28-38, 2017. [Article \(CrossRef Link\)](#)
- [41] Q. Kong, X. Feng, and Y. Li, "Music genre classification using convolutional neural network," in *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, 2014. [Article \(CrossRef Link\)](#)
- [42] D. Jang, S. Shin, J. Lee, and S.-J. Jang, "Web-based platform for music production/playing/distribution," in *Proc. of Int. Workshop on Advanced Image Technology (WAIT)*, 2016. [Article \(CrossRef Link\)](#)
- [43] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp.2825-2830, 2011. [Article \(CrossRef Link\)](#)
- [44] E. Frank, M. A. Hall, and I. H. Witten, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", *Morgan Kaufmann, Fourth Edition*, 2016. [Article \(CrossRef Link\)](#)



**Jongseol Lee** received the B.S. and M.S. degrees in Information & Communication engineering from Chungbuk National University, Korea, in 1996 and 2001. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Konkuk University. Since 2011, he has worked as a senior research engineer at the Smart Media Research Center of Korea Electronics Technology Institute (KETI), Seoul, Korea. His research interests include smart media, audio processing, music analysis, recommendation and retrieval.



**Myeongchun Lee** received received a B.S. degree in Media Information Engineering from Kangnam University, in 2011, a M.S. degree in Computer Science from Yonsei University in 2014. Since 2014, he has worked as a research engineer at the Smart Media Research Center of Korea Electronics Technology Institute (KETI), Seoul, Korea. His research interests include machine learning, deep learning, music information retrieval, natural language processing and data analysis.



**Dalwon Jang** received the B.S., M.S., and Ph.D degrees from Korea Advanced Institute of Science and Technology, in 2002, 2003, and 2010, respectively, all in electrical engineering. Since 2010, he has worked at the Smart Media Research Center of Korea Electronics Technology Institute (KETI), Seoul, Korea. His research interests include content identification, music information retrieval, multimedia analysis, and machine learning.



**Kyoungrro Yoon** received a B.S. degree in Computer and Electronic Engineering from Yonsei University, Seoul, Korea in 1987, an M.S.E. degree in Electrical Engineering/Systems from the University of Michigan, Ann Arbor in 1989, and a Ph.D. in Computer and information Science from Syracuse University in 1999. He was a principal researcher and a group leader in the Mobile Multimedia Research Lab, LG Electronics Institute of Technology from 1999 to 2003. He joined the school of Computer Science and Engineering in 2003 as an assistant professor and became a full professor in 2012. He is with the department of Smart ICT Convergence, since 2017. He served as a co-chair of Ad Hoc Group on User Preferences and the chair of Ad Hoc Group on MPEG Query Format and Ad Hoc Group on MPEG-V of ISO/IEC JTC1 SC29 WG11 (a.k.a. MPEG). He also served as the chair of the Metadata Subgroup and JPSearch Ad Hoc Group of ISO/IEC JTC1 SC29 WG1 (a.k.a. JPEG). He is an editor of various international standards such as ISO IS 15938-12, 23005-2, 23005-5, 23005-6, 24800-3, 24800-5, and 24800-6. His main research interests include smart media system, image processing, multimedia information and metadata processing.