Sciences and Techniques of Information Management
Volume 11, Issue 3, 2025

# Automatic classification of information in Western music and Iranian Traditional Music: A comparative and literature-based analysis

**Amir Vafaeian**
BA in Iranian Music Composition and PhD in Information Science and Knowledge Studies, Kharazmi University, Tehran, Iran.
amirvafa@gmail.com

**Hamid Keshavarz**
Department of Information Science and Knowledge Studies, Faculty of Education and Psychology, Shahid Beheshti University, Tehran, Iran.
ha_keshavarz@sbu.ac.ir

**Ali Salampoor**
MA in Classical Music Performance, University of Tehran, Iran.
a_salampour@yahoo.com

**Brady D. Lund**
Department of Information Science, University of North Texas, Texas, US.
Brady.Lund@unt.edu

**Hamid Reza Saeidnia**
Department of Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran.
hamidrezasaednia@gmail.com

**Yaghoub Norouzi**
Department of Knowledge & Information Science, University of Qom, Qom, Iran.
ynorouzi@gmail.com

## Abstract

**Purpose:** The present article aims to analyze research conducted in the field of automatic music classification, distinguishing the major research patterns in Iranian traditional and Western music classification comparatively for the first time.

**Method**: A literature review was conducted on Western and Iranian traditional music. To locate all relevant publications and documents, appropriate English and Persian keywords were used to search leading Iranian databases such as Noormags, Magiran, Irandoc, Elamjoo, and university library OPACs. English keywords were also applied in global citation databases, including Scopus, Google Scholar, and Web of Science. Relevant terms were identified by referencing various texts on Western, Asian, and Iranian music, as well as sources related to automatic detection.

**Cite this article:** Vafaeian, A., Keshavarz, H., Salampoor, A., Lund, B. D., Saeidnia, H., & Norouzi, Y. (2025 Automatic classification of information in Western music and Iranian Traditional Music: A comparative and literature-based analysis. *Science and Technology of Information Management*, 11 (3), 317-352. https://doi.org/10.22091/stim.2024.10737.2098

**Received:** 2024-05-11 ; **Revised:** 2024-08-18 ; **Accepted:** 2024-08-19 ; **Published online:** 2024-08-29

**Article type:** Research Article **Published by:** University of Qom.

**Findings**: The results showed that the methods used in each of the three main stages of detection–including pre-processing, audio feature extraction, and recognition–differ between Western and Iranian traditional music due to the different attributes and nature of each music genre.

**Conclusion**: This study can aid in the development of algorithms and methods for automatic music identification, contributing new knowledge to the field of music and technology by providing a comprehensive and comparative analysis of scientific works in the field of Automatic Music Classification. Additionally, selecting a method suitable for Iranian traditional music–which ultimately would lead to the automatic detection of melody models and keys–requires a comprehensive and accurate knowledge of the components of this music genre, which is now neglected in related research. The findings of this study can be valuable to researchers, artists, and music enthusiasts, enhancing our understanding of music as an art form.

**Keywords**: Music information retrieval, Automatic music classification, Western music, Iranian traditional music.

**علوم و فنون مدیریت اطلاعات**

دوره ۱۱، شماره ۳، ۱۴۰۴

# طبقه‌بندی خودکارِ اطلاعات در موسیقی غربی و موسیقی سنتی ایران: تحلیلی تطبیقی مبتنی بر متون

**امیر وفائیان**

دکتری علم اطلاعات و دانش شناسی از دانشگاه خوارزمی و کارشناس آهنگسازی ایرانی (نویسنده مسئول). amirvafa@gmail.com

**حمید کشاورز**

دانشگاه شهید بهشتی. ha_keshavarz@sbu.ac.ir

**علی سلام‌پور**

کارشناس ارشد نوازندگی موسیقی کلاسیک غربی. a_salampour@yahoo.com

**بردی لوند**

دانشگاه تگزاس شمالی، دانشکده علوم اطلاعات امریکا. brady.lund@unt.edu

**حمیدرضا سعیدنیا**

دانشگاه تربیت مدرس تهران. hamidrezasaednia@gmail.com

**یعقوب نوروزی**

استاد دانشگاه قم، قم، ایران. ynorouzi@gmail.com

## چکیده

**هدف:** مقاله حاضر با هدف تحلیل پژوهش‌های انجام‌شده در زمینهٔ طبقه‌بندی خودکارِ موسیقی، به تفکیکِ شیوه‌های پژوهشی اصلی در طبقه‌بندی موسیقی سنتی ایرانی و غربی برای نخستین بار به شیوهٔ مقایسه‌ای انجام شد.

**روش:** برای انجام پژوهش، مروری بر ادبیات مرتبط با موسیقی سنتی غربی و ایرانی انجام شد. برای بازیابی منابع، از کلیدواژه‌های انگلیسی و فارسی مرتبط برای جست‌وجو در پایگاه‌های برگزیدهٔ ایرانی مانند نورمگز، مگیران، ایرانداک، و اُپک کتابخانه‌های دانشگاهی استفاده شد. در همین راستا کلماتِ کلیدی انگلیسی نیز در پایگاه‌های استنادی جهانی مانند اسکوپوس، گوگل‌اسکالر و وب‌آوساینس استفاده شد. اصطلاحات مرتبط با ارجاع به متون مختلف در زمینهٔ موسیقی غربی، آسیایی، ایرانی و همچنین منابعی در زمینهٔ تشخیص خودکارِ موسیقی شناسایی شد.

---

**استناد به این مقاله:** وفائیان، الف.، کشاورز، ح.، سلام‌پور، ع.، لوند، ب.، سعیدنیا، ح.، و نوروزی، ی. (۱۴۰۴). طبقه‌بندی خودکارِ اطلاعات در موسیقی غربی و موسیقی سنتی ایران: تحلیلی تطبیقی مبتنی بر متون. علوم و فنون مدیریت اطلاعات، ۱۱(۳)، ۳۵۲-۳۱۷ https://doi.org/ 10.22091/stim.2024.10737.2098

**تاریخ انتشار آنلاین:** ۱٤۰۳/۰٦/۰۸ **تاریخ پذیرش:** ۱٤۰۳/۰۵/۲۹ ؛ **تاریخ اصلاح:** ۱٤۰۳/۰۵/۲۸ ؛ **تاریخ دریافت:** ۱٤۰۳/۰۲/۲۲ ؛

 **نوع مقاله:** پژوهشی **ناشر:** دانشگاه قم

**یافته‌ها:** نتایج نشان داد که روش‌های استفاده‌شده در هر یک از سه مرحلهٔ اصلی تشخیص شامل *پیش‌پردازش*، *استخراجِ ویژگی‌های صوتی* و *مرحلهٔ شناسایی*، میانِ موسیقی سنتی غربی و ایرانی تفاوت وجود دارد که این امر ناشی از تفاوت ماهیتِ بین این دو نوع موسیقی است.

**نتیجه‌گیری:** این مطالعه می‌تواند به توسعهٔ الگوریتم‌ها و انتخاب روش‌های مناسب برای شناسایی خودکارِ موسیقی غربی و ایرانی کمک کرده و دیدگاه جامعی در زمینهٔ طبقه‌بندی خودکار موسیقی ایرانی ارائه دهد. همچنین، این پژوهش می‌تواند در تشخیص خودکارِ گام‌ها و گوشه‌های موسیقی ایرانی، راهنمای خوبی برای پژوهشگران، هنرمندان، و علاقه‌مندان به موسیقی ایرانی باشد؛ این امر به نوبهٔ خود، نیازمند شناخت و بازتعریفِ مفاهیم موسیقی ایرانی است که در پژوهش‌های پیشین مرتبط، به‌طور عمیق بدان پرداخته نشده است.

**کلیدواژه‌ها:** بازیابی اطلاعات موسیقیایی، طبقه‌بندی خودکار موسیقی، موسیقی غربی، موسیقی سنتی ایرانی.

## 1. Introduction

Information retrieval is the science of searching for documents and information, as well as metadata related to these documents. In practice, although there are overlaps and similarities between data recovery for audio and textual information, each has its own theoretical foundation, structure, and technology. The data retrieval process begins when the user submits the query to the system. These queries are matched with the information in the database depending on the application and type of information, i.e. text, image, audio, etc. (Gavahian Jahromi, 2010, p. 30). Many academic and public libraries use metadata retrieval systems to access scientific resources, including books, journals, conference papers, audio and video files, and other documents. Nowadays, people expect more from the virtual world, and among them, music fans want to access the intended piece or work of music online and meet their musical information needs by searching and browsing music informational databases (Aucouturier & Pampalk, 2008).

The field of Music Information Retrieval (hereafter MIR) is one of the subfields of information retrieval, which tries to organize music items, determine the relationships among musical data, or even produce music pieces, and strives to meet the real needs of professional and non-professional music consumers. Musical data include bibliographic data (such as artist's name, genre, and type of music or year), textual data (such as data available on official websites, blogs, or news articles), social data (people who purchase or share an album or monophonic music), or acoustic or musicological information (such as data extracted from audio signals and/or MIDI files) (Bello, 2007).

MIR has an interdisciplinary nature and a community of international researchers in various fields who are developing it continually (Vafaeian, 2018, p. 31). This field is based on sciences such as musicology, mathematics, statistics, sound engineering, computer science, information science, librarianship, cognitive psychology, linguistics, human-information interaction, law, and business (Downie, 2003, p. 296), and each of them considers MIR from a specific aspect. Therefore, this field attempts to solve problems in music and meet the information needs of various music enthusiasts through synergizing and applying knowledge from various fields. Automatic music identification and MIR are also helpful for indexing different genres and national music.

In short, the interdisciplinary field of MIR, that sometimes synonymous with the term "music informatics", is a set of activities to analyze and describe musical or acoustic content by using digital signal processing, artificial intelligence, and other computational methods. "Music technology," which goes beyond the field of MIR, is considered a more general term, and research areas such as musical interaction, performance modeling, and music cognition fall into this category (Heydarian, 2016, p. 35).

Automatic music analysis and processing have been the focus of the research field MIR for nearly half a century. The term MIR was first used by Michael Kassler (1966) in the journal Perspectives of New Music. The MIR acronym was the title of a programming language for music developed under the direction of Tobias Robison in collaboration with Hubert, Howe and Kassler during a research project at Princeton University (Kassler 1966: 59). The International Society for Music Information Retrieval (ISMIR) has been holding annual meetings and conferences on music-related data since 1999, which is one of the most important events in this field (Heydarian 2016: 34-35). With the availability of more musical data in recent decades, MIR researchers have tried to create methods to automatically extract meaningful descriptors from audio files – a subfield known as "automatic music classification."

Most of the existing research on automatic music classification (AMC) has focused on Western music and genres. In general, these research studies can be divided into six main categories: genre classification, musical instruments, mood classification, similarity in musical pieces or non-original music, chord detection, and scale/tonality/melody detection. However, research in the field of MIR in oriental music in general and Iranian traditional music, in particular, is far sparser than Western music, both in terms of the number and the variety of research methodologies.

The present work is an attempt to review the research conducted in the field of AMC in Western and Iranian classical music exhaustively and comparatively, which is done for the first time. Therefore, in the following sections, some research in the field of AMC related to Western classical music is presented, and then, by systematic search in the relevant literature, the position of the research on automatic retrieval of Iranian classical music information is analyzed. It should be noted that, except for studies related to the classification and identification of music genres in Western music, only classical or traditional music of the countries is considered, and other genres of music (such as pop, jazz, local, and regional music, etc.) have not been included in this research.

Carrying out such a comparative study can reveal the differences and similarities in automatic identification techniques of Western and Eastern music based on the analysis of patterns, features, and structures present in Western and Eastern music. This study can contribute to the development of algorithms and methods for automatic music identification and offer new insights in the field of music and technology that can lead to an improvement in our understanding and comprehension of this art. Therefore, the findings of this research can be valuable for researchers, artists, and music enthusiasts in this regard.

## 2. Automatic Classification of Western music

Some of the research on MIR are related to classifying, clustering, modeling, extracting monophonic and polyphonic musical features, pattern matching and identifying similarities among audio tracks (Gavahian Jahromi 2010: 30). The types of classifications in this area also include a wide range of topics such as identifying the artist (i.e. musician/singer), recognizing the origin of the music or the mood that the music creates, automatically separating and recognizing the instruments used in orchestral performances, music note recognition and automatic key/mode/scale recognition in a musical piece (Vafaeian 2017: 31). Accordingly, the classification and automatic detection of music also seek to help this field by analyzing the theoretical foundations of Western music on the one hand and the application of machine learning techniques like Support Vector Machine (SVM) on the other.

The number of studies conducted in the field of AMC in Western music is considerable in a way that they can be classified into six main categories as follows: automatic genre classification, musical instrument recognition, automatic music mood/emotion classification of tracks, similar or non-original musical works, chord identification, and melody/scale/key/tonality detection. This categorization is presented for the first time based on a literature review into automatic identification of Western music by the authors.

### 2-1. Automatic Genre Classification

Extensive research has been done on music retrieval in areas such as instrument recognition, rhythm extraction, music tracks and so forth (Habibi Aghdam and Homayounpour 2010: 33). In the meantime, the number of the research on MIR confirms that music genre classification is more popular than the other topics to the point that more than half of the research published in the journal International Association for the Music Information Retrieval is published on this topic. Most of the current music genre classification techniques uses machine learning techniques (Vishnupriya, and Meenakshi, 2018). On the other hand, Feature Extraction is the most crucial task for audio analysis. Tzanetakis (2015: 64) argues that one of the practical applications of audio features extraction is the classification of different genres of music, as one of the most common research methods of most case studies.

Automatic retrieval of music genres includes finding tracks that belong to a particular genre. As a result, the growing research related to automatic classification of Western music based on content has led researchers to consider exploring such features as the number of words uttered per minute, the number of words per line of the lyric, or signs such as semicolons, tempo and rhythm of the music piece, or using different audio features to distinguish different genres of Western music such as pop, jazz and traditional. For example, Xu et al. (2003) classified different genres of Western music using the SVM learning method and features like spectral coefficient, zero-crossing rate and Mel frequency coefficient (Abdollahzadegan, Jafari and Dirand 2014: 918).

At present, there are challenges such as ineffective categorization, low precision, and time-intensive processes in current automated music genre classification methods and Lina (2022) propose a novel approach using deep belief networks and sparse representation for more effective genre classification. Another proposed method for classifying the genre of Western music is to recognize the type of certain instruments used in the music track. For example, the electric guitar is used more in rock music than in Western traditional music. Also, different genres of Western music can be classified within a certain range of rhythm and tempo of the music piece and therefore they can be distinguished from each other. Aiming at the problems of poor classification effect, low accuracy, and long time in the current automatic classification methods of music genres, an automatic classification method of music genres based on deep belief network and sparse representation is proposed.

Some of the conducted research on automatic classification of music genres are: Ndou, Ajoodha and Jadhav (2021), Cheng et al. (2021), Ceylan, Hardalaç, Kara and Hardalaç (2021), Kikuchi, Aoki, and Dobashi (2020), Henrique Foleiss, and Fernandes Tavares (2020), Vishnupriya, and Meenakshi (2018), Rosner, and Kostek (2018), Aguiar, Costa, and Silla (2018), Karkavitsas and Tsihrintzis (2012), Baniya, Lee and Li (2014), Ariyaratne and Zhang (2012), Aryafar, Jafarpour and Shokoufandeh (2012), Chen, Chen and Truong (2012), Yang et al. (2011), Seyerlehner, Widmer and Knees (2010), Völkel et al. (2010), Simsekli (2010), Silla and Koerich (2010).



## 2-2. Automatic Mood Classification of Music Tracks

Emotions play a major role in terms of our preferences, choices, and decision-making. The mood is the state of emotion perceived by a music listener while listening to a music track. However, these two words are used interchangeably in the literature. In other words, the music we listen to, the emotions it influences, and the resulting mood are all interrelated (Chaturvedi et al. 2022). Music automatically enters into daily human life and expresses emotions, synchronizing the listener's emotions. The current mental state of a person is linked to music, and recognizing human emotions based on sound, rhythm, harmony, melody, and timing is a complex process.

One of the emerging domains in MIR over the past 20 years has been the mood classification of music. Automatic music mood classification based on emotion in music tracks could be divided into several general categories: epic, romantic, happy, sad, scary, mellow, upbeat, aggressive, and so on. Some common machine learning techniques for music mood classification are multilayer perceptron (MLP), support-vector machine (SVM), and Naïve Bayes.

For example, Krishnaiah, A. & Divakarachari, P. B. (2021) proposed hybrid spectral feature extraction technique extracted and combined the spectral features of audios such as Spread, centroid, skewness, Mel Frequency Cepstrum Coefficients (MFCCs) and Linear Predictor Coefficients (LPCs) reduced the dimensionality complexity enhanced the success rate of the Multi-Support Vector Machine (MSVM) classifier for Indian mood classification.

Chaudhary, D., Singh, N.P. & Singh, S. (2019) believe that the classification models used for predicting emotions were not very efficient. They propose a novel approach named "hashtag graph generation" for automatic emotion detection. The method consists of two steps: a training process and a testing process. It is compared with support vector machines, the k-nearest neighbor approach, and a convolutional neural network in terms of accuracy, precision, recall, specificity, F-measure, geometric mean, root mean square error (RMSE), and computational cost. This technique achieves the best performance across all evaluation parameters.

Some of the research in mood classification would include Han, D. et al. (2022), Chaturvedi et al. (2022), and Rajan et al. (2021). Das, Satpathy, and Debbarma (2021), Mo and Niu (2019), Bhattarai and Lee (2019), Hu, Choi and Downie (2017), Ren, Wu and Roger Jang (2015), Mokhsin et al. (2014), Sturm (2012), Ujlambkar and Attar (2012), Panagakis and Kotropoulos (2011), Laurier, (2011), Sun and Tang (2009) and Skowronek, McKinney and Van De Par (2006).



## 2-3. Recognition of Musical Instruments

One of the most important challenges in describing music is the accurate recognition of instruments within audio data. Automatic music classification serves as the foundation for various advanced AI applications in the musical domain, such as automatic multi-instrument classification and information extraction from polyphonic audio (Mahanta, Basisth, Halder, et al., 2023). One of the earliest comprehensive studies on the classification and recognition of musical instruments was conducted by Herrera-Boyer, Peeters, and Dubnov (2003). This research identified two practical applications of musical instrument classification: "labeling" monophonic recordings, sound samples within sample libraries, or new patches created with a given synthesizer; and "indexing," which involves recognizing the main instruments in a musical mixture (for example, identifying a saxophone solo in the middle of a song). According to this study, solving the first problem (labeling) is easier than addressing the second (indexing).

The logical approach to recognizing instruments is the initial separation of sounds based on different audio sources and then segmenting and classifying these tracks. In other words, research on classification has focused more on isolated sounds, assuming that the separation and segmentation have already been performed. This implies the use of a sound sample collection (usually isolated notes) consisting of different classes and instrument families (Herrera-Boyer, Peeters, and Dubnov 2003: 4).

Once the audio segmentation is complete, the tracks need to be labeled. Two different families of algorithms can be used for assigning labels: If the labels are known in advance, pattern recognition, discrimination, or supervised learning techniques are appropriate choices. However, when the labels are unknown and must be inferred from the data, unsupervised learning or clustering techniques are more suitable (ibid., p. 3).

As Liu and Wan (2001) mentioned, among a wide range of features extracted, the best feature set was selected by the Sequential Forward Feature Selection (SFS) method, and then the instruments were classified. Firstly, 58 features were extracted for each record in a music database, containing 351 audio files. Then, the extracted features were normalised by the mean and standard deviation. The best feature set was selected for high-precision classification, and then a new feature was created from a combination of previous features to minimise the amount of classification error. This process continued until all features were selected. The length of the audio files varied from 0.1s to about 10s, and each audio file was divided into 256 sample windows, with half of the adjacent windows overlapping. Compared to the comprehensive search approach requiring a lot of computation time, the SFS software can be used to create an optimised feature set in less time.



In the most recent studies for music automatic classification (such as Ke, Lin, and Sharma, 2021; Koszewski and Kostek, 2020; Solanki and Pandey, 2019; Gururani, Sharma, and Lerch, 201,9, and Jeyalakshmi, Murugeshwari, and Karthick, 2018), Deep Neural Network (DNN) and Deep Convolutional Network (CNN) are the most common classifiers; others are HMM, SVM, and KNN. Feature extraction algorithms, such as MFCC, PL, P, and RASTA-PLP, ZC, R, are used before this step. There are also larger datasets, such as OpenMIC are used for polyphonic instrument recognition.

One of the rare research studies on the classification of Iranian musical instruments was conducted by Mousavi, M. H., and Prasath, V. B. S. (2019). In this study, features such as Mel-Frequency Cepstral Coefficients (MFCCs), Spectral Centroid, Zero Crossing Rate (ZCR), and Entropy Energy were utilized. For feature selection, the Fuzzy Entropy measure was applied, and classification was performed using a Multi-Layer Neural Network (MLNN). These features were extracted from audio signals in a novel database containing samples from seven Persian musical instrument classes: Ney, Tar, Santur, Kamancheh, Tonbak, Ud, and Setar.

## 2-4. Automatic Detection of Similarity or Non-Original Music

For those who intend to perform previously published music tracks, it is necessary to obtain permission from the original producers (i.e., composers and musicians) to avoid copyright infringement. Therefore, distinguishing original works from non-original ones is a key concern for musicians and the music industry.

On the other hand, due to the large volume of music tracks, manual detection may be difficult or even impossible. Therefore, automatic identification and separation of the two groups (i.e., original pieces and non-original ones) has been explored by researchers in the field of Music Information Retrieval (MIR). Automatic identification of cover songs is useful for a wide range of applications, from organizations involved in licensing copyrighted music to fans seeking new versions of their favorite songs (Bertin-Mahieux and Ellis, 2011).

The term a non-original piece of music or a previously recorded song or track. In other words, a cover song is a new recording or performance of a song originally recorded by a different artist (Chen et al., 2015).

Many approaches have been proposed for this task (e.g., Serra, Gomez, & Herrera, 2010, pp. 108–311). A may differ from the original in terms of key, structure, timbre, tempo, arrangement, or vocal language. Consequently, cover song identification has recently attracted considerable attention. However, most state-of-the-art methods rely on similarity search, which requires a large number of similarity computations to retrieve potential cover versions for a query recording (Silva, de Souza, & Batista, 2015).

## 2-5. Automatic Chord Recognition (ACR)

A chord is any harmonic set of frequencies consisting of multiple notes (also called "pitches") that are heard as if sounding simultaneously. In other words, the simultaneous sounding of three, four, five, six, or seven different notes is called a chord, from which the triad chord is the simplest and most common type of chord. In this type of chord, a triad is created by basing each of the notes of a scale as the onset point (root) and by adding the other two notes called the third and fifth of the chord (Fleser 2010: 21-22). There are four types of triad chords: major, minor, augmented, and diminished. Automatic chord detection is an important task in the analysis of music transcription and can contribute to applications such as structural segmentation and key detection.

The first research on the field of Automatic Chord Recognition (ACR) was done by Fujishima in 1999. Since then, the field has evolved significantly from early knowledge-based systems towards data-driven methods, with neural networks (Pauwels, O'Hanlon, Gómez, Sandler, 2019) and deep learning methods. With the advances of machine learning technologies, data-driven feature extraction and sequence modeling approaches are being widely explored for automatic chord recognition tasks (Wu & Li, 2019).

Using chromagram, Harte and Sandler (2005) presented an algorithm that calculates the Constant-Q transformation that leads to automatic chord recognition. In this study, a chromogram consisting of 36 binaries was used to determine the exact location of the semitones. Then, each of the three adjacent bins was merged to form a 12-bin semitonic chromogram, and then compared with 48 triad chord patterns. The chromogram that had the most energy is assigned to the corresponding chord. The overall accuracy of the results was 53.9% to 77.9%.

Sheh & Ellis (2003) also developed an algorithm for recognizing chords using the chroma feature (i.e., Pitch Class Profiles: PCP) and Markov's hidden model. In this model, the data are trained using the Expectation Maximisation (EM) algorithm. The recognition accuracy on 20 tracks of the Beatles database was 75%. Moreover, this study showed that for recognizing chords, the chroma feature is better than the cepstral coefficients.



Some other recent studies in ACR are: Hernández, Guerrero, and Macías-Díaz, 2022; Mabpa et al., 2021; Pauwels, O'Hanlon, Gómez, Sandler, 2019; and Wu & Li, 2019.

### 2-6. Automatic Detection of Scale, Tonality and Melody Similarity

To recognize the scale (or key) of a music track, it is first necessary to identify the pitch set of the track. In addition to identifying the tonic note (the most stable pitch), it is also required to determine the hierarchical relationship of other pitches in the pitch set (c 2002: 18). Noland (2009: 23-24) divided key recognition methods of the Western music into three general categories:

- *The method based on music theory*: In this case, patterns based on diatonic-scale degrees (major and minor) are created. Thus, based on the chromatic scale, the notes that are used in this pattern are assigned 1 and other notes are assigned 0.
- *The method based on cognitive studies*: In this method, instead of assigning 0 or 1 for notes (tones), the patterns are weighted based on the importance of each note.
- *The method based on collecting statistics of real music*: For doing so, diatonic intervals are determined and weighted based on real data and the music tracks themselves and not only based on the theoretical fundamentals of the music (quoted by Heydarian 2016: 36).

Noland (2009) proposed an algorithm for estimating and recognizing the tonality of a music track using signal processing and a hidden Markov model. This study demonstrated that reducing the sampling frequency by a factor of 16 decreases computational cost and generally slightly improves scale detection results. Noland (2009, p. 3) also noted that there is no single ideal set of parameters applicable to all music genres; however, by adjusting parameters for each genre, more accurate results can be achieved.

Some studies, such as Chew (2002), have focused on scale changes and modulation. Although Chew acknowledged the difficulty of identifying the boundary points where modulation occurs, he addressed this challenge by introducing a novel geometric model for detecting tonality, called the spiral array, along with a boundary search algorithm. The spiral array is a spatial representation of classes of pitches, chords, and scales that serves as the foundation for recognizing the tonality of musical pieces (Chew 2002, p. 19).

Izmirli (2006: 127-132) suggests two models for finding the key (scale): a model based on correlation and a model based on tonality representation with smaller dimensions. The first model is based on confidence-weighted correlation while the second one is the first model's independently confirmed performance to be used as a benchmark to evaluate the second model. The results of this study showed that the best scale recognition performance remains almost constant from 12 dimensions down to 6, which means a tonal representation with at least 6 dimensions can be used to capture the essential portion of information.

In general, the automatic recognition process in Western music can be divided into three general steps: pre-processing, audio features and recognition. The pre-processing stage includes two sub-stages: framing signals and data normalisation. The second stage includes three sub-stages of extraction, selection and creating the feature vector. The last step is usually done using one of the machine learning or pattern recognition techniques (Figure 1).

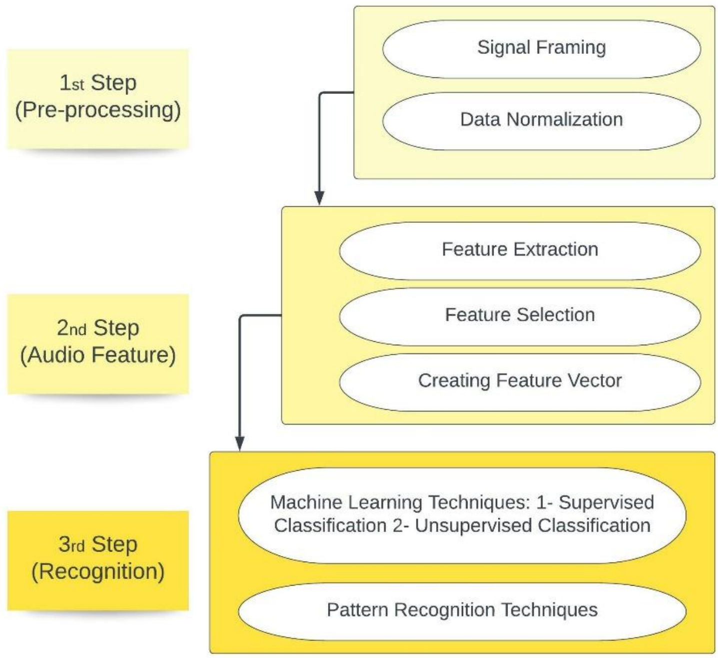


**Figure 1. The main steps of automatic detection in Western music**

## 3. Automatic Recognition of Iranian Traditional Music: A Literature Review



Iranian music could be categorised under the broader family of music, which is called “modal music” in international musicology. Closer members of this family in Eastern music are Arab and Turkish. In the advanced or traditional performance form of this family, the *Avaz* (songs) or *Mugams* are combined to form a set of "multimode" systems in Iranian traditional music that is called “*Dastgāh*” (Talai 2015: 11).

Broadly speaking, the mode is defined as the underlying skeleton of the structure of the *Dastgāhs*/*Avazes* and the *Dastgāh* is a multi-mode cycle consisting of melodies of modes organised in a cyclical pattern. *Avaz* can also be considered as sub-keys or its derivative that is almost similar to the *Dastgāh* in terms of model status. However, *Avaz* has a simpler structure than *Dastgāh* (Asadi 2004: 46, 51; Talai 2015: 11).

Importantly, the term "*Dastgāh*" in Iranian traditional music is somehow equivalent to the "Diatonic Scale or Key/scale" in Western music, "*Raga*/*Raag*" in Indian, "*Mugam*" in Turkish and "*Maqam*" in

Arabic music. There are seven main *Dastgāhs* in Iranian classical/traditional music including *Mahour, Shur, Sehgah, Chahargah, Homayoun, Nava* and *Rastpanjgah*. A sequence of eight consecutive notes/tones that have a ratio of certain intervals from each other is called a diatonic scale in Western music. To simplify the concept of the *Dastgāh* in Iranian music, the same definition is sometimes used in which each of the seven *Dastgāhs* differ in scales on a same octave.

Iranian traditional music is a delicate and profound art form with a rich cultural heritage. Three key characteristics define this music: 1) the extensive Iranian repertoire (*Radif*) and its melodic framework (*Gusheh*), 2) the unique complexities and intricacies in the interval patterns across seven scales, and 3) the diverse and subtle performance techniques. These features contribute to the challenges in learning and teaching this music, particularly due to its reliance on memory, which can lead to ambiguity. Therefore, the automatic identification of Iranian traditional music would greatly assist learners, performers, enthusiasts, and students with varying levels of musical knowledge by simplifying the learning process and enhancing its quality. Ultimately, this advancement would help promote Iranian traditional music among diverse social groups and within the international music community.

### 3-1. Literature Search Strategy

To locate all relevant publications and documents in both English and Persian, keywords were searched in key Iranian databases, including Noormags, Magiran, Irandoc, Elamjoo, and the Online Public Access Catalogues (OPAC) of university libraries. Additionally, English keywords were used to search global citation databases such as Scopus, Google Scholar, and Web of Science. The keywords were derived by consulting various texts on Western, Asian, and Iranian music, as well as retrieved sources in the field of automatic detection. Ultimately, 15 sources related to the automatic detection of Iranian traditional music were identified and analysed.

The total number of studies conducted specifically to classify the keys (*Dastgāhs*) and melody models (*Gushehs*) of Iranian traditional music was 15, one of which (Heydarian 1999) only dealt with the issue of determining the base frequency of a note in an Iranian musical instrument and has not considered key and melody model detection. Out of the 15 studies conducted, only three were published between 1999 and 2010 (Heydarian 1999; Darabi 2003; Gavahian Jahromi 2010), while the remaining 12 were published from 2011 to 2019 (see Table 1). This indicates a significant increase in attention to this issue in recent years. Examining the list of sources and their publication dates reveals that, except for three studies, the rest were conducted independently and concurrently. More importantly, due to either a lack of awareness or limited access to related and similar sources, these studies did not reference previous work. Additionally, of the 15 studies, six were written in English, while the remainder were in Persian.

Of these 15 research projects, 10 were dissertations (1 undergraduate, 7 master's, and 2 doctoral) and 5 were scientific articles (3 conference papers and 2 journal papers). Over time, the desire to publish works in the form of articles has increased. Also, the main specialty of the 10 authors of these works was electronic engineering. The scientific disciplines and majors of the other authors included computer engineering, sound engineering, mechanical engineering, mathematical sciences, biomedicine, and information science. Except for Vafaeian (2019), who attempted to automatically identify and distinguish between melody models (*Gushehs*), the other researchers focused on recognizing the main keys (*Dastgāhs*) in Iranian traditional music. Based on the data collected, most of these researchers lacked sufficient mastery of the theoretical foundations of Iranian traditional music, including the significant differences between *Gushehs* and *Dastgāhs*. Consequently, their understanding of the key, melody model, and repertoire in this genre has often been inadequate.

**Table 1. A detailed overview of the works retrieved from the literature review**

| | Row | 1 | 2 |
|---|---|---|---|
| Thematic knowledge | Author (year) | Heydarian (1999) | Darabi (2003) |
| | Genre | Master's Dissertation (Tarbiat Modarres University) | Bachelor's Dissertation (Khaje Nasir |
| | Academic discipline | Electronic Engineering | Electronic Engineering |
| Methodological knowledge | Sampling frequency (Fs) (kHz) | 8 | – |
| | Frequency window length (milliseconds) and other audio file features | mono, bit rate 16<br>Nf=2048<br>Rectangular window | 100 ms<br>(wav, 8 bits) |
| | Length of track cuts (second) | 1 s | – |
| | Number of Evaluated music tracks | 130 notes | – |
| | Instrument / Song | Glockenspiel | – |
| | Data collection conditions | Recording notes | – |
| | Evaluated keys/Dastgahs | – | – |
| | Training method (classifier) | – | – |
| | Evaluated audio features | Pitch | Pitch<br>(obtaining FFT and extracting repeated notes of the piece) |
| Results | Accuracy of key detection | Identification of notes with 92.96% accuracy | – |

| | | |
|---|---|---|
| Results | Accuracy of key detection | Average overall accuracy: 71%<br>Setar: 64.51<br>Band playing: 66.67<br>Santur: 81.81<br>Ney: 90<br>*Camancheh*: 64.7<br>Violin: 54.54 |
| Methodological knowledge | Evaluated audio features | Pitch Profile |
| | Training method (classifier) | 1. Minkowski distance (rank 4)<br>2. Calculating the degree of similarity between the pitch profile and 4 scale patterns |
| | Evaluated keys/Dastgahs | 4 main Dastgahs |
| | Data collection conditions | CD and cassette tape<br>(from reputable music companies) |
| | Instrument / Song | *Tar*: 55 tracks<br>*Setar*: 31<br>Band playing: 26<br>Santur: 21<br>*Ney*: 20<br>Fiddle: 15<br>Violin: 11 |
| | Number of Evaluated music tracks | 179 tracks |
| | Length of track cuts (second) | 20-25 s |
| | Frequency window length (milliseconds) and other audio file features | 512 ms (wav, hamming, 16 bins, single-channel) |
| | Sampling frequency (Fs) (kHz) | 44.1 |
| Thematic knowledge | Academic discipline | Electronic Engineering |
| | Genre | Master's Dissertation (Shahrood University of Technology) |
| | Author (year) | Gavahian Jahromi (2010) |
| | Row | 3 |

| | | | |
|---|---|---|---|
| Results | Accuracy of key detection | Mahour detection with more than 70% accuracy | Average error rate –Neural network: 87.10 - Minimum distance: 05.14 |
| Methodological knowledge | Evaluated audio features | Pitch (Top 20 peaks in each window) | Pitch obtaining FFT and extracting 5 repetitive notes |
| | Training method (classifier) | RBF Artificial Neural Network (Supervised Learning) | - neural network - minimum distance |
| | Evaluated keys/Dastgahs | Mahour detection from 5 Dastgahs/ other Avazes | Shur Dastgah and *Afshari* Avaz |
| | Data collection conditions | Recording in a radio studio (by a musician with a certain tune) | - Album of Santur Saba repertoire |
| | Instrument / Song | Setar | Dulcimer |
| | Number of Evaluated music tracks | 195 tracks | 80 |
| | Length of track cuts (second) | 5-10 s | 60 s |
| | Frequency window length (milliseconds) and other audio file features | wav, mono, bit rate 16 | 100 ms |
| | Sampling frequency (Fs) (kHz) | 44.1 | 44.1 |
| Thematic knowledge | Academic discipline | Audio Engineering | Electronic |
| | Genre | Master's Dissertation (Department of Broadcasting, IRIB | Master's |
| | Author (year) | Mahmoodan (2012) | Bayat (2013) |
| | Row | 4 | 5 |

| | | |
|---|---|---|
| Results | Accuracy of key detection | Identified devices with an overall accuracy of 93%<br>- Neural network: 61.20<br>–KNN۷۳/۳۶ :<br>–Naïve Bayes۸۴/۳۲ : |
| Methodological knowledge | Evaluated audio features | Pitch<br>(Extracting notes and determining their frequency intervals) |
| | Training method (classifier) | MLP neural network (two middle layers and five neurons in each layer) –KNN<br>–Naïve Bayes |
| | Evaluated keys/Dastgahs | 7 Dastgahs |
| | Data collection conditions | Recording tracks<br>(played by music teachers with tuning fork) |
| | Instrument / Song | Tar and dulcimer |
| | Number of Evaluated music tracks | 46 tracks<br>(42 tracks with tar and 4 tracks with dulcimer) |
| | Length of track cuts (second) | Complete performance of the track (whole track) |
| | Frequency window length (milliseconds) and other audio file features | The window length varies and is selected based on the onset of the attack phase mp3 |
| | Sampling frequency (Fs) (kHz) | 44.1 |
| Thematic knowledge | Academic discipline | Electronic Engineering |
| | Genre | Master's Dissertation (Shiraz University) + conference paper (2014: Ferdowsi University) |
| | Author (year) | Abdollahzadegan (2014) |
| Row | | 6 |

| | Row | 7 | 8 |
|---|---|---|---|
| Results | Accuracy of key detection | Average total accuracy: 74.5% | Average error rate 28.2% |
| Methodological knowledge | Evaluated audio features | - MFCC<br>- ZCR<br>- Spectral Rolloff<br>- Spectral Flux | –Pitch<br>obtaining FFT and extracting repetitive notes |
| | Training method (classifier) | HMM (Hidden Markov Model)<br>(to find the optimal model: Baum-Welch algorithm and to evaluate the system performance: 10-fold cross-validation) | Single-layer artificial neural network with the |
| | Evaluated keys/Dastgahs | 7 Dastgahs | Homayoun Dastgah |
| | Data collection conditions | *Haft Dastgāh* training album, played by Majid Kiani (Sarv Setah Cultural Institute) | – |
| | Instrument / Song | Dulcimer | – |
| | Number of Evaluated music tracks | 330 tracks (from 180 melody models) | 70 tracks (30 melody models/melody in Homayoun) |
| | Length of track cuts (second) | 30 s | – |
| | Frequency window length (milliseconds) and other audio file features | 30 ms<br>wav | – |
| | Sampling frequency (Fs) (kHz) | 32 | – |
| Thematic knowledge | Academic discipline | Computer engineering (Artificial Intelligence) | Electronics |
| | Genre | Master's Dissertation (Semnan University) | Conference Paper |
| | Author (year) | Peivandi (2015) | Darabi, Azimi and |

| | | | |
|---|---|---|---|
| Results | Accuracy of key detection | - Detection of keys with an overall accuracy of 85%<br>-The most similarity between the two keys of Mahour and Chahargah (85%)<br>- The least similarity between the two keys of Segah and Chahargah (43%) | 100% accuracy in detection of all keys! |
| Methodological knowledge | Evaluated audio features | Pitch | Pitch |
| | Training method (classifier) | Fuzzy logic type 2 | Artificial neural network (three layers) |
| | Evaluated keys/Dastgahs | 5 Dastgahs | 5 Dastgahs |
| | Data collection conditions | CD and cassette tape | Not specified |
| | Instrument / Song | Singing (three outstanding masters) and a small number of dulcimers, strings, trumpets and fiddles | Not specified |
| | Number of Evaluated music tracks | 210 melody models | Not specified |
| | Length of track cuts (second) | About 60 s | Not specified |
| | Frequency window length (milliseconds) and other audio file features | 45 ms | Not specified |
| | Sampling frequency (Fs) (kHz) | Each track was different from the other | Not specified |
| Thematic knowledge | Academic discipline | Electronic and computer engineering | Electronic and computer |
| | Genre | Conference Paper (AISP) | Conference Paper (AISP) |
| | Author (year) | Abdoli (2011) | Hajimolahoseini, Amirfattahi, & |
| | Row | 9 | 10 |

| | | | |
|---|---|---|---|
| Results | Accuracy of key detection | The superiority of the RBF kernel function in the SVM classifier over other methods. Detection accuracy in RBF method 65 to 95% for different keys | Violin: 72%<br>Ney: 65<br>Singing: 56% |
| Methodological knowledge | Evaluated audio features | -Pitch<br>- MFCC<br>– Inharmonicity<br>- Spectral Centroid | Pitch<br>(top 20 or maximum local peaks in each window modeled from Mahmoudan research) |
| | Training method (classifier) | Main classifier: SVM<br>Other methods<br>KNN and MLP | Three-layer artificial neural network of MLP types, 30 neurons in hidden layer (supervised learning) |
| | Evaluated keys/Dastgahs | (12 Dastgahs/song + Kurdish Bayat song) | 7 Dastgahs |
| | Data collection conditions | Taken from five CDs | - Album of old singers<br>- Recording for Ney and violin |
| | Instrument / Song | Tar and Setar<br>(Four unidentified musicians) | Ney, violin and song |
| | Number of Evaluated music tracks | 250 melody models<br>from the repertoire of Mirza Abdullah | 348 tracks |
| | Length of track cuts (second) | 20 s | 10-20 s |
| | Frequency window length (milliseconds) and other audio file features | 3 s | wav, bit rate 8 |
| | Sampling frequency (Fs) (kHz) | 44.1 | 44.1 |
| Thematic knowledge | Academic discipline | Electronics Engineering and Biomedicine | Mechanical engineering |
| | Genre | Journal article | Mechanical engineering |
| | Author (year) | Abbasi Layegh, Haghipour, c Sarem (2013) | Beigzadeh & Belali Koochesfahani |
| Row | | 11 | 12 |

| | | | |
|---|---|---|---|
| Results | Accuracy of key detection | _ | Total average accuracy: 86.21% |
| Methodological knowledge | Evaluated audio features | - Pitch Histogram<br>- Chroma<br>- Spectral Average | Temporal and spectral |
| | Training method (classifier) | - Manhattan Distance (original)<br>- GMM<br>- Cross-Correlation | Deep neural network |
| | Evaluated keys/Dastgahs | 5 Dastgahs | 7 Dastgahs |
| | Data collection conditions | - Recording tracks (for dulcimer) | _ |
| | Instrument / Song | Santur 11 and 12 Vault (original)<br>- A small number for fiddle, Ney and ensemble instruments) | Violin and Ney |
| | Number of Evaluated music tracks | - Santur: 104<br>Band playing: 22<br>- Fiddle: 11<br>- Piano: 7 | 1137 tracks<br>631 Ney<br>506 violin |
| | Length of track cuts (second) | _ | 16 s |
| | Frequency window length (milliseconds) and other audio file features | 92.9 ms<br>hann window<br>not-overlapping | 510 ms |
| | Sampling frequency (Fs) (kHz) | _ | 8192 |
| Thematic knowledge | Academic discipline | Electronics engineering, music | Computer Engineering (Artificial |
| | Genre | Ph.D. Thesis (London Metropolitan University) | Bachelor's Dissertation (Khajeh |
| | Author (year) | Heydarian (2016) | Rezazadeh Azar, et al. (2019) |
| | Row | 13 | 14 |

| | | |
|---|---|---|
| | Row | 15 |
| Thematic knowledge | Author (year) | Vafaeian (2019) |
| | Genre | Ph.D. Thesis (Kharazmi University) |
| | Academic discipline | Information science, music |
| Methodological knowledge | Sampling frequency (Fs) (kHz) | 44.1 |
| | Frequency window length (milliseconds) and other audio file features | - 64 ms<br>- Studio environment - single-channel - m4a format |
| | Length of track cuts (second) | Whole track |
| | Number of Evaluated music tracks | Tar: 38<br>Setar: 31<br>Santur: 73<br>*Barbat*: 31 |
| | Instrument / Song | The six main melody models of Shur from Mirza Abdullah repertoire |
| | Data collection conditions | Recording tracks |
| | Evaluated keys/Dastgahs | 6 melody models in Shur Dastgah<br>(Daramad-e-aval, Owi, Shahnaz, Mogadameh-e-Qarcheh, Qarcheh and |
| | Training method (classifier) | LDAand<br>Fisher scale |
| | Evaluated audio features | Temporal, Spectral, Cepstral features<br>–Pitch<br>(extracting notes and determining their frequency intervals) |
| Results | Accuracy of key detection | Tar: 82.11<br>Setar: 91.66<br>Santur: 86.66<br>The average accuracy of detecting melody models in the second stage (%)<br>Note recognition accuracy: 87.9% |



## 3-2. Results and Discussion

In an independent study employing the documentary-library research method, all 15 studies were reviewed and compared across various aspects, including the research framework and scope, type of database, primary sources for the music basics section, presented suggestions, as well as notable strengths and weaknesses. The analysis revealed that research on the automatic recognition of Iranian traditional musical instruments is generally fragmented, insufficient, and incomplete. These studies have not achieved the desired outcomes due to the absence or limited access to a comprehensive database of Iranian traditional music, the lack of integration among these studies, and, most importantly, the absence of a comprehensive, accurate, and specialized perspective on the practical and theoretical foundations of Iranian traditional music (Vafaeian et al., 2018).

Mahmoodan and Banooshi (2012: 6) argue that among the three features of pitch (frequency intervals), rhythm, and resonance, only pitch plays a role in determining Iranian musical instruments. Therefore, the most important processing step for raw data is to obtain the spectral frequency using the fast Fourier transform (FFT) and identify the dominant frequencies in the track. It appears that most research conducted in the initial stage of key detection—specifically, the detection and recognition of notes—has relied on this same feature (Mahmoodan 2012; c 2013; Abdollahzadegan 2014; Darabi, Azimi, and Nojumi 2006; Abdoli 2011; Hajimolahoseini, Amirfattahi, and Zekri 2012; Beigzadeh and Belali Koochesfahani 2016). Bayat (2013) identified Navak (frequency pitch) as the primary feature for detecting notes used in a melody model or track, thereby facilitating the automatic classification of music. After defining the top peaks as frequency pitches, he constructed the feature vector based on five repetitive notes of a melody mode.

In Mahmoodan (2012) and Beigzadeh and Belali Koochesfahani (2016), the feature vector is based on the top 20 peaks in each window. In addition to the pitch feature, some studies have incorporated other features in automatic key detection. For example, Abbasi Layegh, Haghipour, and Najafi Sarem (2013) used three additional features alongside pitch: Mel-frequency Cepstral Coefficients, inharmonicity components, and spectral centroid. Heydarian (2016) examined three features: pitch histogram, chroma, and spectral mean. Gavahian Jahromi (2010) utilized the pitch profile feature. Additionally, Peivandi (2015), in research on the automatic classification of Western music genres, employed four features—Mel-frequency Cepstral Coefficients, zero-crossing rate, spectral roll-off, and spectral flux—to recognize Iranian musical instruments; however, no satisfactory results were obtained (Table 1, row 7).

Moreover, Beigzadeh & Belali Koochesfahani (2016) demonstrated that some crucial information for recognizing Iranian musical instruments can be extracted from the spectral frequency of an input track (signal). However, there are additional features and factors beyond the frequency content of the music track; neglecting or failing to extract these would increase errors in detecting and classifying the key (Beigzadeh & Belali Koochesfahani 2016, p. 108). In other words, despite the significant importance of frequency and pitch features in automatic detection, it is also necessary to consider the sequence and chronological order of the notes alongside the spectral frequency to accurately detect the key and melody model (Beigzadeh and Belali Koochesfahani 2016, p. 116).

In the first part of his research, Vafaeian (2019) examined the importance of each of the 21 temporal, spectral, and cepstral features in detecting and distinguishing the main melody models of the Shur key. For doing this, first, using Fisher's scale, each of the 21 audio features is scored in MATLAB software, and then, from among them, the top three features are selected. Then, using the linear discriminant analysis category, the possibility of separating the melody models from each other was examined for each instrument separately and in general. The results showed that among the 21 extracted features, the cepstral features obtained the highest score in separating the melody models from each other and were in a better state than the temporal and spectral features. However, none of these features were able to distinguish and detect the six main melody models of the Shur key. In the second part of this research, first, the note recognition stage was performed on the whole music track. Afterward, by defining the melody path and creating a database of a note-time matrix for each representative (Moarref) phrase and then matching and examining its similarity with the note-time matrix of each music track, the possibility of automatic detection of the main melody models of the Shur key was examined (Table 1, row 15).

In some of these studies, note detection has been done in the first stage. Subsequently, to separate the keys from each other, different decision-making and classification methods have been used, and different types of artificial neural networks have received more attention than other methods (8 works out of 15 works). Therefore, both in the automatic detection of Iranian traditional musical instruments and in Western music, the modelling and automatic detection of music using neural networks are mostly used as a classifier and a method of teaching systems, compared to other methods. Other types of methods used in this research include Gaussian Mixture Models, SVM, different measurement intervals (estimation based on minimum distance estimation, Manhattan distance and Minkowski space), K-nearest neighbour classifier, hidden Markov model, linear discriminant analysis, and fuzzy logic.

Although the number of studies conducted in the field of automatic classification and detection of music outside Iranian and Western music is significant, their function could be divided into six general groups (Figure 2): automatic classification of music genre, recognition of musical instruments, automatic detection of the mood of musical tracks, automatic detection of a chord, detection of melody to find a specific music track (e.g. Track ID and Sound Hound) and automatic detection of scale (major or minor). Moreover, some research has been conducted to extract the rhythm of music tracks. Studies related to the classification of music genres have gained a larger share over the past 20 years.

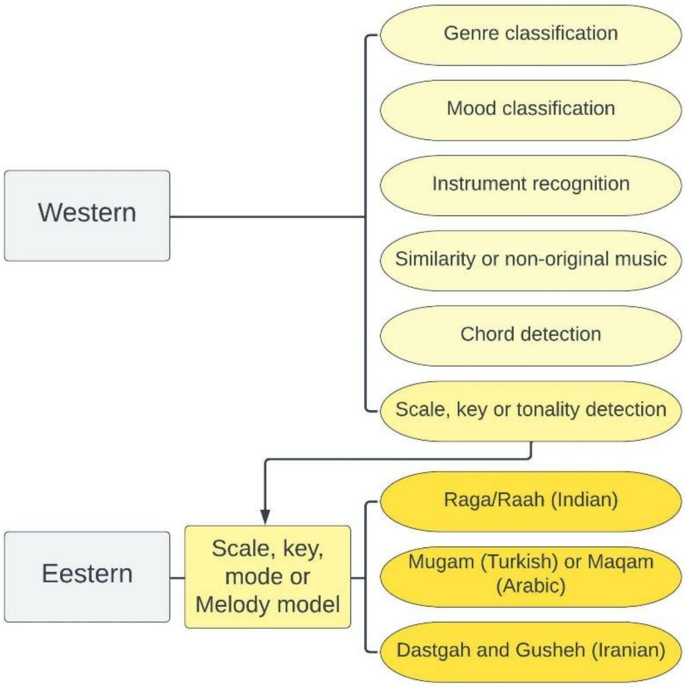


**Figure 2. Automatic music detection in Western and Eastern music**

However, almost all studies carried out in the field of automatic detection of Oriental music, including Iranian traditional music, have been in the field of scale detection (with different titles such as Key, Mugam, or Raga) as if other functions such as the classification of genre, instrument, mood and chords have not been among the concerns of Iranian music researchers for various reasons that need to be evaluated separately.

All 15 studies have been conducted by Iranian researchers, indicating that there has been little demand for such research among non-Iranian scholars. When examining sources and research on Iranian traditional music from the 1950s to the 1970s, we find a significant number of contributions by non-Iranian researchers. However, following the Iranian Revolution in 1979, the number of such sources declined markedly. In other words, after this pivotal historical event, the interest in theorizing Iranian traditional music and conducting related research diminished significantly. Consequently, the lack of interest among non-Iranian researchers can be extended to the field of automatic recognition of Iranian music. Most research on automatic key detection in Iranian music has followed a trajectory similar to the three stages observed in Western music, as illustrated in Figure 1.

In addition, an extensive attempt was made to find similar or identical methods. However, the methods used in each of the three main stages of automatic music recognition would be different between Western and Iranian traditional music due to the different attributes and nature of each music genre. In other words, regarding the automatic detection of melody models and keys, it is not possible to use the same methods used in the automatic detection of Western music without considering the features of Iranian traditional music. The inefficient and undesirable results of most Iranian music research also confirm this finding.

Therefore, some supervised machine learning algorithms, such as multilayer perceptron, k-nearest neighbor classifier, and SVM, commonly used in the classification of Western music genres, would not have good results for automatically detecting keys and melody models in Iranian music. For example, Peivandi (2015) treated the automatic detection of keys (*Dastgāh*) in Iranian music as analogous to the classification of music genres in Western music by using audio feature extraction. However, the concept of genre in Iranian traditional music fundamentally differs from that in Western traditional music, as all melody models and repertoire systems of Iranian music can be encompassed within a single genre called "Iranian traditional music.

As Noland (2009: 3) argued, there is no definitive or ideal set of parameters applicable to all music genres, and the accuracy of research findings can be improved by tailoring parameters to each specific genre. In most studies related to the automatic detection of Western music, the selection of sound features varies depending on the requirements and applications of automatic music detection. However, this approach has not been applied to Iranian traditional music. Due to the unique characteristics of Iranian traditional music—including representative phrases in each melody model, specific musical intervals within each key and melody model, the presence of microtones in scales and modes, Mugam, and tonic and dominant notes specific to each melody model—it is not feasible to fully implement the data elements, metadata, and methods used for the automatic organization of Western traditional music in the context of Iranian traditional music. Furthermore, research conducted thus far on the automatic detection of Iranian traditional music, except for a few recent studies, has not yielded satisfactory results; only the main keys of Iranian traditional music have been detected, while the melody models remain unidentified.

Mahmoodan and Banooshi (2012: 6) believe that among the three features of pitch frequency (notes), rhythm, and resonance, only the pitch frequency feature plays a role in determining Iranian musical instruments, which does not seem to be true. Nevertheless, there are features and factors other than the frequency content of the music track, which, if ignored or not extracted, would lead to increased errors in the detection and classification of the key. In other words, despite the importance of the frequency pitch in the automatic detection of Iranian music, all the required information cannot be summarised in this feature.

**5. Conclusion**

To automatically detect the key and melody model in Iranian music, it is essential to consider the sequential and chronological features of note performance or rhythm in music tracks, alongside the spectral frequency. These two elements—time and pitch frequency—together form the melodic progression of a music track. Except for Vafaeian (2019), who attempted to detect the melodic progression of melody models and tracks by constructing a net-time matrix of representative phrases, thereby enabling the separation of melody models, time and rhythm have generally been overlooked in research on the automatic detection of Iranian music. Although this topic has attracted researchers' attention for over a decade, the automatic detection and classification of keys in Iranian traditional music still require more extensive and precise efforts. Selecting a method suitable for Iranian traditional music that ultimately facilitates the automatic detection of melody models and keys demands comprehensive and accurate knowledge of the components of this musical genre.

## References


Abbasi Layegh, M., Haghipour, S. & Najafi Sarem, Y. (2013). Classification of the *Radif of Mirza Abdollah* a canonic repertoire of Persian music using SVM method. *Gazi University Journal of Science, Part A: Engineering and Innovation*, 1 (4), 57-66

Abdoli S. (2011). Iranian traditional music *Dastgāh* classification, *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 275-280

Abdollahzadegan, S. (2014). *Automatic classification of Iranian traditional music based on scale (Dastgāh)*. Master's Dissertation, Department of Electrical Engineering, Shiraz University, Shiraz. [In Persian]

Abdollahzadegan, S., Jafari, S. & Dirand, M. (2014). Automatic detection of scale (*Dastgāh*) of Iranian traditional music based on Recital Performance of Tar and Santur by intelligent extraction of note. *20th National Annual Conference of Iranian Computer Society*. Ferdowsi University, Mashhad. [In Persian]

Aguiar, R. L., Costa, Y. & Silla, C. N. (2018). Exploring Data Augmentation to Improve Music Genre Classification with ConvNets, *2018 International Joint Conference on Neural Networks (IJCNN), 1-8*

Ariyaratne, H.B. & Zhang, D. (2012). A Novel Automatic Hierarchical Approach to Music Genre Classification. *International Conference on Multimedia & Expo (ICME) Workshops 2012*, 564-569

Aryafar, K., Jafarpour, S. & Shokoufandeh, A. (2012). Automatic musical genre classification using sparsity-eager support vector machines. *International Conference on Pattern Recognition*, 1526-1529

Asadi, H. (2004). Theoretical foundations of Persian classical music: *Dastgāh* as a multi-modal cycle. *Mahour Music Quarterly*, 6 (22), 43-56. [In Persian]

Aucouturier, J. J. & Pampalk, E. (2008). Introduction–From Genres to Tags: A Little Epistemology of Music Information Retrieval Research. *Journal of New Music Research*, 37 (2), 87-92

Baniya, B.K., Lee, J. & Li, Z.N. (2014). Audio feature reduction and analysis for automatic music genre classification. Systems, Man and Cybernetics (SMC), *IEEE International Conference*, 457-462

Bayat, H. (2013). *Classification and Identification of scale (Dastgāh) in Iranian music*. Supervisor: Hamed Sajedi. Master's Dissertation, Department of Electronic and Electrical Engineering, Shahed University, Tehran. [In Persian]

Beigzadeh, B. & Belali Koochesfahani, M. (2016). Classification of Iranian traditional musical modes (*Dastgāh*) with artificial neural network. *Journal of Theoretical and Applied Vibration and Acoustics*, 2, (2), 107-118

Bello, JP. (2007). Audio-Based Cover Song Retrieval Using Approximate Chord Sequences: Testing Shifts, Gaps, Swaps and Beats. *Proceedings of the 7th International Conference on Music Information Retrieval, (ISMIR)*, Victoria, Canada, 239-244.

Bertin-Mahieux, T. & Ellis, D.P. (2011). Large-scale cover song recognition using hashed Chroma landmarks. *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 117*-120.

Bhattarai, B. & Lee, J. (2019). Automatic Music Mood Detection Using Transfer Learning and Multilayer Perceptron, *International Journal of Fuzzy Logic and Intelligent Systems; 19*(2), 88-96.

Ceylan, H.C, Hardalaç, N., Kara, A. and Hardalaç F. (2021). Automatic Music Genre Classification and Its Relation with Music Education, *World Journal of Education, 11*(2), 36-45.

Chaturvedi, V., *et al.* (2022). Music mood and human emotion recognition based on physiological signals: a systematic review. *Multimedia Systems, 28,* 21–44.

Chaudhary, D., Singh, N.P. & Singh, S. (2019). Automatic music emotion classification using hashtag graph. *International Journal of Speech Technology, 22,* 551–56. https://doi.org/10.1007/s10772-019-09629-2

Chen, N., et al. (2015). Modified Perceptual Linear Prediction Liftered Cepstrum (MPLPLC) Model for Pop Cover Song Recognition. *ISMIR*, 598-604.

Chen, S.-H., Chen, S.-Hu. & Truong, T.-K. (2012). Automatic music genre classification based on wavelet package transform and best basis algorithm. *International Symposium on Circuits and Systems (ISCAS 2012)*, 3202-3205

Cheng, Y. H., Chang, P. C., Nguyen, D. M., & Kuo, C. N. (2020). Automatic Music Genre Classification Based on CRNN. *Engineering Letters*, *29*(1).

Chew, E. (2002). The spiral array: an algorithm for determining key boundaries. *Music and artificial intelligence*, *2445*,18-31.

Darabi, N. (2003). *Producing and analyzing music digital signals: designing Nava scale programming language to produce music pieces and automatic recognition of scales (Dastgāhs)* for Persian music, Bachelor's Dissertation, Department of Electrical Engineering, Khaje Nasir Toosi University of Technology, Tehran. [In Persian]

Darabi, N., Azimi, N. H., & Nojumi, H. (2006). Recognition of *Dastgāh* and Maqam for Persian Music with Detecting Skeletal Melodic Models, *The second annual IEEE BENELUX/DSP Valley Signal Processing Symposium (SPS-DARTS 2006),* Belgium, March. http://sharif.edu/~nojumi/pubfiles/rdmpmdsmm.pdf

Das, S., Satpathy, S., & Debbarma, S. (2021). Multimodal Music Mood Classification Framework for Kokborok Music. *Progress in Advanced Computing and Intelligent Engineering, Vol 1299, 163–175,* Springer, Singapore.

Downie, J.S. (2003). *Annual Review of Information Science and Technology, Chapter 7: Music Information Retrieval.* 295-340.

Fleser J. (2010). *The chord wheel: The ultimate tool for all musicians*. Tran A, Jahanshahi. Tasnif Publications. [In Persian].

Fujishima, T. (1999). Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music. *ICMC Proceedings 1999, 464-467*

Gavahian Jahromi, E. (2010). *Persian Music Classification using Pitch Profile Feature*. Supervisor: Hossein Marvi. Advisor: Ali Soleimani. Master's Dissertation, Faculty of Electrical and Robotic Engineering, Shahrood University of Technology, Shahrood. [In Persian].

Gururani, S., Sharma, M. and Lerch, A. (2019). An Attention Mechanism for Musical Instrument Recognition. *20th International Society for Music Information Retrieval Conference (ISMIR), Delft, the Netherlands.*

Habibi Aghdam, H., & Homayounpour, M. M. (2010). Automatic recognition of music genre. *Signal and Data Processing Journal*, 7(1), 33-52. [In Persian].

Hajimolahoseini, H., Amirfattahi, R., & Zekri, M. (2012). Real-time Classification of Persian Musical *Dastgāhs* Using Artificial Neural Network. *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012),* 157-160

Han, D., et al. (2022). A survey of music emotion recognition. *Frontiers of Computer Science, 16,* Article number: 166335

Harte, C. A., & Sandler, M. B. (2005). Automatic chord identification using a quantised chromogram. *Proceedings of th 118th Convention of the Audio Engineering Society,* Marcelona, May, 1-5.

Henrique Foleiss, J., & Fernandes Tavares, T. (2020). Texture selection for automatic music genre classification, *Applied Soft Computing, 89,* 106127.

Hernández, R., Guerrero, A., & Macías-Díaz, J.E. (2022). A template-based algorithm by geometric means for the automatic and efficient recognition of music chords. *Journal of Evolutionary Intelligence. 17*(1), 467-481.

Herrera-Boyer, P., Peeters, G., & Dubnov, S. (2003). Automatic Classification of Musical Instrument Sounds. *Journal of New Music Research, 32*(1), 3-21. http://www.tandfonline.com/doi/abs/10.1076/jnmr.32.1.3.16798

Heydarian, P. (1999). *Music note recognition for Santur*. Master's Dissertation, Department of Electrical Engineering, Tarbiat Modarres University, Tehran. [In Persian].

Heydarian, P. (2016). *Automatic recognition of Persian musical modes in audio musical signals* (Doctoral dissertation, London Metropolitan University).

http://portal.acm.org/citation.cfm?doid=1859799.1859815

Hu, X. Choi, K., & Downie, S. (2017). A framework for evaluating multimodal music mood classification, *JASISIT, 68(2), 273-285*

Izmirli, O. (2006). Audio key finding using low-dimensional spaces. *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, 8-12 October, Victoria, Canada, 127-132.

Jeyalakshmi, C., Murugeshwari, B., & Karthick, M. (2018). HMM and K-NN based Automatic Musical Instrument Recognition, *2nd International Conference on I-SMAC, 350-355.*

Karkavitsas, G. V., & Tsihrintzis, G. A. 2012. Optimization of an Automatic Music Genre Classification System via Hyper-Entities. *Proceedings of the 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 449-452.

Kassler, M. (1966). Toward Musical Information Retrieval. *Perspectives of New Music, 4* (2), 59-67.

Ke, J., Lin, R., & Sharma, A. (2021). An Automatic Instrument Recognition Approach Based on Deep Convolutional Neural Network. *Recent Advances in Electrical & Electronic Engineering, 14(6), 660-670*

Kikuchi, Y., Aoki, N., & Dobashi, Y. (2020). A Study on Automatic Music Genre Classification Based on the Summarization of Music Data, *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 705-708.*

Koszewski, D., & Kostek, B. (2020). Musical Instrument Tagging Using Data Augmentation and Effective Noisy Data Processing, *Journal of the Audio Engineering Society*, *68*(1/2), 57-65

Krishnaiah, A., & Divakarachari, P. B. (2021). Automatic Music Mood Classification using Multi-class Support Vector Machine based on Hybrid Spectral Features. *International. Journal of Intelligent Engineering & Systems, 14* (5), 102.

Laurier, F. C. (2011). *Automatic Classification of Musical Mood by Content-Based Analysis.* Ph.D. Thesis, Universitat Pompeu Fabra.

Liu, M., & Wan, C. (2001). Feature selection for automatic classification of musical instrument sounds. *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries – JCDL,* 247-248.

Mabpa, P. et al. (2021). Automatic Chord Recognition Technique for a Music Visualizer Application, *9th International Electrical Engineering Congress (iEECON), 416-419*

Mahanta, S.K., Basisth, N.J., Halder, E. et al. (2023). Exploiting cepstral coefficients and CNN for efficient musical instrument classification. *Evolving Systems*. https://doi.org/10.1007/s12530-023-09540-x

Mahmoodan, S. (2012). *Automatic classification of audio signals using artificial neural network*. Master's Dissertation in Audio Engineering, Department of Broadcasting, IRIB University, Tehran. [In Persian].

Mahmoodan, S., & Banooshi, A. 2012. *Automatic classification of Mahoor scale (Dastgāh) using artificial neural network.* 2nd Conference of Iranian Society of Acoustics and Vibrations (*ISAV*). Sharif University, Tehran. [In Persian].

Mo, S. and Niu, J. (2019). A Novel Method Based on OMPGW Method for Feature Extraction in Automatic Music Mood Classification, *IEEE Transactions on Affective Computing*, *10*(3), 313-324.

Mokhsin, M. B., Rosli, N. B., Wan Adnan, W. A., & Abdul Manaf, N. (2014). Automatic music emotion classification using artificial neural network based on vocal and instrumental sound timbres. In *New Trends in Software Methodologies, Tools and Techniques* (pp. 3-14). IOS Press.

Mousavi, M. H., & Prasath, V. B. S. (2019). Persian Classical Music Instrument Recognition (PCMIR) Using a Novel Persian Music Database. *9th International Conference on Computer and Knowledge Engineering (ICCKE), 122-130.*

Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches, *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 1-6.*

Noland, K. C. (2009). *Computational tonality estimation: signal processing and hidden Markov models*. PhD Thesis, Queen Mary, University of London.

Pan, L. (2022). [Retracted] Automatic Classification Method of Music Genres Based on Deep Belief Network and Sparse Representation. *Journal of Mathematics*, *2022*(1), 8752217. https://doi.org/10.1155/2022/8752217

Panagakis, Y., & Kotropoulos, C. (2011). Automatic music mood classification via low-rank representation. *European Signal Processing Conference*, (Eusipco), 689-693.

Pauwels, J., O'Hanlon, K., Gómez, E., & Sandler, M. (2019). 20 Years of Automatic Chord Recognition from Audio. *Proceedings of the 20th Conference of the International Society for Music Information Retrieval (ISMIR).*

Peivandi, J. (2015). *Design an efficient System to Detect Traditional Iranian Music style*. Supervisor: Hadi Soltanizadeh. Master's Dissertation in Artificial Intelligence, Faculty of Electrical and Computer Engineerin, Semnan University, Semnan. [In Persian].

Rajan, R., Antony, J., Joseph, R. A., & Thomas, J. M. (2021, November). Audio-mood classification using acoustic-textual feature fusion. In *2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS)* (pp. 1-6). IEEE.

Ren, J.M., Wu, M.J., & Roger Jang, J.S. (2015). Automatic Music Mood Classification Based on Timbre and Modulation Features, *IEEE Transactions on Affective Computing, 6* (3), 236-246. doi:10.1109/TAFFC.2015.242783

RezezadehAzar, S., Ahmadi, A., Malekzadeh, S., & Samami, M. (2018). Instrument-Independent Dastgah Recognition of Iranian Classical Music Using AzarNet. *arXiv preprint arXiv:1812.07017*.

Rosner, A., & Kostek, B. (2018). Automatic music genre classification based on musical instrument track separation. *Journal of Intelligent Information Systems, 50, 363–384*

Serra, J., Gomez, E., & Herrera, P. (2010). Audio Cover Song Identification and Similarity: Background, Approaches, Evaluation and Beyond. In Zbigniew W. Ra´s and Alicja A. Wieczorkowska (Eds.). *Advances in Music Information Retrieval*, (308-332). Springer.

Seyerlehner, K., Widmer, G., & Knees, P. (2010). A Comparison of Human, Automatic and Collaborative Music Genre Classification and User Centric Evaluation of Genre Classification Systems. *Adaptive Multimedia Retrieval,* 118-131.

Sheh, A., & Ellis, D. P. W. (2003). Chord segmentation and recognition using EM-trained Hidden Markov Models. *Proceedings of the 4thInternational Conference on Music Information Retrieval (ISMIR),* Baltimore,1-7.

Silla, C. N. & Koerich, A. (2010). Improving automatic music genre classification with hybrid content-based feature vectors. *Proceedings of the 2010 ACM Symposium on Applied Computing*, 1702-1707. Retrieved from http://dl.acm.org/citation.cfm?id=1774453

Silva, D.F., de Souza, V.M. and Batista, G.E. (2015). Music Shapelets for Fast Cover Song Recognition. *ISMIR, 1 October 2015, 441-447.*

Simsekli, U. 2010. Automatic Music Genre Classification Using Bass Lines. *International Conference on Pattern Recognition (ICPR),* 4137-4140.

Skowronek, J., McKinney, M., & Van De Par, S. (2006). Ground Truth for Automatic Music Mood Classification. *Proceedings of the 7th International Conference of Music Information Retrieval (ISMIR 2006), 2,* 395-396. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.7371&rep=rep1&type=pdf

Solanki, A., & Pandey, S. (2019). Music instrument recognition using deep convolutional neural networks. *International Journal of Information Technology, 14* (3),1-10.

Sturm, B. L. (2012). Making Explicit the Formalism Underlying Evaluation in Music Information Retrieval Research: A Look at the MIREX Automatic Mood Classification Task. *Computer Music Multidisciplinary Research (CMMR),* 89-104.

Sun, X., & tang, Y. (2009). Automatic Music Emotion Classification Using a New Classification Algorithm, *Second International Symposium on Computational Intelligence and Design*, 540-542

Talai, D. (2015). *Radif analysis: based on the Notation of Mirza Abdollah's Radif with Annotated visual description*. Ney Publishing Co. [In Persian].

Tzanetakis, G. (2015). *Music Information Retrieval (Courseware).* http://marsyas.cs.uvic.ca/mirBook/mirBook_jan05_2015.pdf

Ujlambkar, A. M., & Attar, V. Z. (2012). Automatic Mood Classification Model for Indian Popular Music. In *2012 Sixth Asia Modelling Symposium* (pp. 7-12). IEEE.

Vafaeian, A. (2017). Study on scientific products in music information retrieval in Scopus. *Scientometrics Research Journal, Winter and Summer, 3* (5): 30-47. [In Persian].

Vafaeian, A. (2019). *Automatic Detection of Shur Dastgāh in the Iranian Traditional Music based on Recital Performance of Tar, Setar Santur and Barbat Instruments*. PhD Thesis, Kharazmi

University, Faculty of Psychology and Education, Department of Knowledge and Information Science. [In Persian].

Vafaeian, A., Borna, K., Sajedi, H., Alimohammadi, D., & Sarai, P. (2018). Automatic Identification and Classification of the Iranian Traditional Music Scales (Dastgāh) and Melody Models (Gusheh): Analytical and Comparative Review on Conducted Research. *Human Information Interaction*, *5*(2), 46-72.

Vishnupriya, S., & Meenakshi, K. (*2018*). Automatic Music Genre Classification using Convolution Neural Network, In *International Conference on Computer Communication and Informatics (ICCCI)*, (pp.1-4).

Völkel, T., Abeßer, J., Dittmar, C., & Großmann, H. (2010, September). Automatic genre classification of latin american music using characteristic rhythmic patterns. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound* (pp. 1-7).

Wu, Y. , & Li, W. (2019). Automatic Audio Chord Recognition with MIDI-Trained Deep Feature and BLSTM-CRF Sequence Decoding Model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(2), 355-366*

Xu, C., Maddage, N. C., Shao, X., Cao, F., & Tian, Q. (2003, April). Musical genre classification using support vector machines. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03).* (Vol. 5, pp. V-429). IEEE.

Yang, X., Chen, Q., Zhou, S., & Wang, X. (2011). Deep Belief Networks for Automatic Music Genre Classification. In *INTERSPEECH* (pp. 2433-2436).