Soft Computing (2021) 25:12249–12260
https://doi.org/10.1007/s00500-021-05898-9

FOCUS

Emotion-based music recommendation and classification using machine learning with IoT Framework

Mohammad Tabrez Quasim[1] · Eman H. Alkhammash[2] · Mohammad Ayoub Khan[1] · Myriam Hadjouni[3]

Accepted: 19 May 2021 / Published online: 15 June 2021

Abstract

Technological advances integrating emotional maturity with established IoT systems are being examined with the emergence of the fourth industrial revolution. In this article, researchers propose an emotion-based music recommendation and classification framework (EMRCF) categorizing songs with high precision following individuals' interpersonal team with memory and emotional songs. In specific, when adding new tunes to an IoT app fortune, methods must be developed that immediately categorize the characters based on people's emotions . That's one of the essential questions for project management. The empathic framework is used to research to identify emotional information. Musical characteristics can be derived from discussions in a micro-enterprise with the task force. Correlation analysis and supporting neural network is used to perform dynamic designation. The innovative prediction accuracy proposed recognizes most of the emotional responses triggered by music audience members and effectively categorizes songs. Furthermore, a comparison study is made with proposed algorithms such as decision trees, deep cognitive system and neighbor-closest, and relevance vector machines. The EMRCF reaches the prediction accuracy of 96.12% and the precision rate of 96.69%, which is not achieved by existing approaches.

Keywords Music recommendation · Machine learning · Internet of things · Emotion · Music classification · Emotion recognition

1 Introduction to the music recommendation system

ICT has assumed the fourth industrial rebellion leadership producing new and innovative product payment systems that never occurred previously (Ayata et al. 2018). It is based on a variety of telecommunications. Since it recognizes a critical technology that maximizes comfort in individuals' everyday lives and provides excellent tools for distinct ways of life and climates, IoT's alterations will bring to the globalized world an immense passion (Seo and Huh 2019).

With increased availability for advanced IoT innovation and detailed analysis, many scientists from the business sector, academic research and funding agencies focus on emotionally intelligent advancement to identify people's feelings and ideas. Empathy evaluation (empathy assessment) is a subject inspired by an instinctive attempt to know human emotions and impeding personal information interpretation (Abdul et al. 2018). Even Snapchat, Gmail, Google and Microsoft, the largest IT corporations, build social technology to obtain or acquire feelings or take over customers and the firm in this study method to incorporate facial reactions (Andjelkovic et al. 2019).

Songs are one of the particular investigation areas for this empathy assessment. As much online material exposure has grown, entertainment has become easily accessible and heard in our daily situations. Nevertheless, per the feelings or feelings, the game's required language or

Communicated by Vicente Garcia Diaz.

✉ Mohammad Ayoub Khan
ayoub.khan@ieee.org

[1] Department of Computer Science, College of Computing and Information Technology, University of Bisha, Bisha, Kingdom of Saudi Arabia

[2] Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Kingdom of Saudi Arabia

[3] Department of Computer Science, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Kingdom of Saudi Arabia

subculture can vary exceptionally suggested by Wang et al. (Wang et al. 2018a). Therefore, innovations are launched to enhance the simplicity of use that analyzes human emotions to recommend songs in a given environment. Since headphones from the AI equipment have been currently introduced and commercialized with Amazon Echo, Smart Speakers, Search Assistance and Virtual Assistants, songs per the consumer emotions can be selected suggested by Chen et al. (Chen et al. 2019). It will gradually broaden offerings relevant to music, such as recommending good music by evaluating people's feelings. Companies can use it to reach clients in different markets like advertising, schooling, media, board games and wellness.

Current human geography investigations have shown that song gives its users an emotional reaction (Yang et al. 2018). The musical taste was heavily associated with character traits and emotions. In parts of the brain where feelings and temperament are addressed, the song's meter, tempo and intensity are controlled.

There are undoubtedly many various influences such as sexuality, maturity, community, personal interests, mood and meaning that rely on the consumer's emotional reaction to a song segment (Wang et al. 2018b). These independent factors enable people to categorize the songs reliably as pleased, sorry, excited or relaxing. Studies examined in recommending programs focused on sentiment relies on two principal dimensions, audio and video (Zangerle et al. 2018). As verbal communication is recognized, researchers isolate audio functions and analyze contemporary British and Scottish songs to translate these functionalities into four simple feelings. The grouping of automated music using such mood groups produces good outcomes .

The most historical and common way to communicate thoughts, feelings and perceptions is body language. Researchers categorize body language into four types: pleasant, depressed, furious and negative, for this article's benefit (Kim and Lim 2018). This study's general intent is to create an economically efficient music app that makes a sensitivity-conscious soundtrack-dependent on the participant's psychological response. While hearing the music, the event process is vibrant. The soundtrack is an effective tool to influence the conscious experience encapsulated, and however, music is such an effective specific cultural platform to enhance the participant's psychological response. This software has a minimum machine resource consumption. The framework for sentiment decides the participant's mood (Chen et al. 2019).

The Music Identification system extracts significant and necessary sound from a record. The suggestion module blends emotional effects with the musical rating framework to suggest the customer (Wang et al. 2018b). The precision and efficiency of this framework were considerably higher than current structures.

The paper's main contribution includes the design of the EMRCF, evaluation of the EMRCF for recommendation of music, obtaining experimentation results and respective justification.

The rest of the research work as follows. Section 2 deals with the background and the literature survey of the music recommendation system. The proposed Emotion-based music recommendation and classification framework (EMRCF) is designed and implemented in Sect. 3, and the software analysis and performance evaluation are discussed in Sect. 4. The conclusion and future scope are illustrated in Sect. 5.

2 Background to the music recommendation system

Yang et al. reported that face emotions could be classified into seven main areas: furious, disgusting, joyful, frightening, surprising, sad and damaging (Yang et al. 2018). In those other terms, for all classes, socioeconomic strata and age groups, these physical features have been internationally identical and accepted by various backgrounds. Suppose nevertheless, people tend to disguise their gestures. In that case, it will not be sufficient to diagnose feelings using only facial features definitively. Using physical cues is a more accurate tool for tracking and recognizing human emotions and intellectual-specific functionality than facial gestures (Schedl et al. 2018). This emotional drawback was resolved using different sensors, including photosynthesis, blood pressure, electrochemical skin sensitivity/viscosity, etc.

Information is driven, or shared screening approaches are used by conventional recommendations that do not understand user sentiment. The use of human feeling with suggestion systems can, furthermore, improve efficiency in suggestion motors. Krämer et al. introduced an online music classification model for stress relief, and the process utilizes PPG fingers form, portable and compact (Krämer 2018). Guo et al. suggested the phone music framework offers context-aware, nano-sensor-based (Guo et al. 2019).

Volokhin et al. described a music technique recommending the participant's pulse and inclination (Volokhin and Agichtein 2018). Li et al. have introduced a custom song suggestion framework with chosen characteristics, meaning and playing experience (Li et al. 2020). Kim et al. proposed an emotionally intensive music suggestion method, known as strengthened sensation metrics (ESM), which connects an orthography feeling metric and a user-based conductivity meter (Kim and Lee 2018). The sensations of consumers are drawn from the phrases shared on social networking, and the song suggestion scheme is carried out through a low-complication interface for smartphones. Researchers suggest a song suggestion method that considers the emotion of participants in their recommendations (Lu and Tintarev 2018).

The technologies guidelines are primarily focused on two considerations: the participant's previous tastes and the potential impact on the suggested music's consumer sentiment (Murthy and Koolagudi 2018). The two methods are relevant to the research. Participant's preference is determined through suggestions based on history, and the present mood is predicted by the framework based on the emotions. The machine senses the perspective of the participant and measures suggestions before and during an album. The Platform employs GSR and PPG to monitor the psychological shifts of the consumer continually (Rachman et al. 2018). Researchers suggested a method of emotion detection dependent solely on GSR indications before the operation. In this analysis, researchers enhance messages with PPG and provide an approach to analyzing different emotional datasets for trains with musical advice (Hsu et al. 2018). The suggested system seeks to improve efficiency by understanding the emotional consumer status of a music recommender system.

Research methods were suggested for the classification of the participant's feelings and actions. Domínguez-Jiménez et al. concentrated on using emotional tissue growth to detect action units formed with temporary and temporary physical expressions (Domínguez-Jiménez et al. 2020). The development of methodologies has increased Convolutionary Neural Networks (CNNs) for speech feelings perception.

Lyrical interpretation is used to classify songs. Although it is comparatively simpler to apply this cryptocurrency approach, it cannot correctly distinguish songs alone. The verbal communication that limits registration to a given program is another apparent problem of this approach (Er and Aydilek 2019). Musical elements such as speed, rhythm and pace to identify the songs' sentiments are another approach for classifying the harmonious atmosphere. This approach consists of removing a range of attributes to identify a confident attitude's trends using such function variables.

In brief, all the existing methods recommend the music with analysis of either by the taste or mood. The proposed methods try to focus on both aspects.

3 Proposed Emotion-based music recommendation and classification framework (EMRCF)

The machine learning and Internet of things-based suggestion framework is part of a feedback suggestion, and the Mel Frequency Coefficient function values are quantified for the information.

Figure 1 shows the architecture of the proposed EMRCF framework. Firstly, researchers selected six categories of songs that comprise the simple database: joyful, silent, emotional, depressed, inspired and enthusiastic. In this document, researchers retrieve the Mel Frequency Coefficient function's magnitude, cluster it and get the soundtrack's worth. Then researchers place in the repository information, including the artist and the description of the song and their function score. If the person enters the suggested sound, the same functionality attribute is derived, comparing the song's length in the collection. If the subjective area is close, all pieces are content-like: the source song's suggestion.

3.1 Element extraction

Mel Frequency Coefficient is an innovation for voice extraction. The following moves are:

Figure 2 shows the workflow of the proposed EMRCF framework. The steps of the framework are explained below.

(1) Interception: The album usually takes 3-five minutes, and specific repeats of material are always present. The estimate is very high and pointless; the function worth of the entire track is enormous. The writers suggested that the encoding entity should only be selected thirty seconds from the music center. The route is chosen for review for just thirty seconds in this article.

(2) Pre-emption: The high-frequency vibration maximum is illuminated. The pre-emphasis converter variable audio signals enhance the increased component to correct the significant proportion of the deleted audio signals. The buffer pre-focus is described in Eq. (1), and ρ is the factor pre-focus function.

$$S_n^{'} = S_n - \rho \times S_{n-1} \quad (1)$$

The buffer focus function is denoted as S_n. The previously obtained variable is denoted as S_{n-1}. The focus factor is denoted as ρ. The predicted function for the next iteration is characterized as $S_n^{'}$.

Figure 3 shows the pictorial representation of $S_n^{'}$. The buffer focus function is denoted as S_n. The

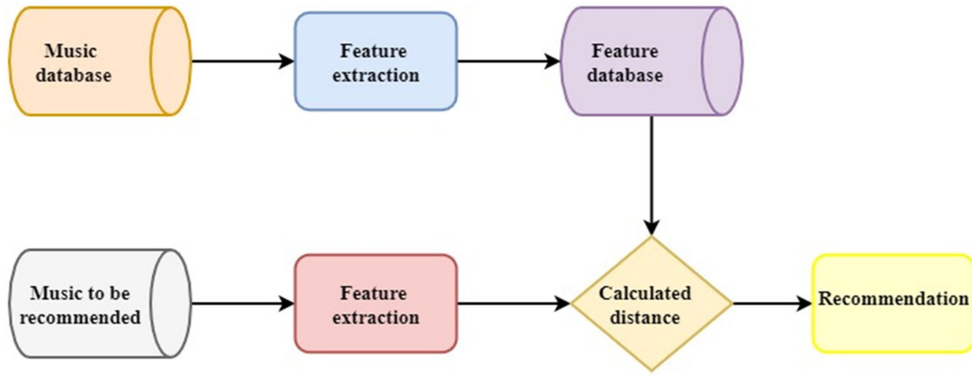


Fig. 1 The architecture of the proposed EMRCF framework

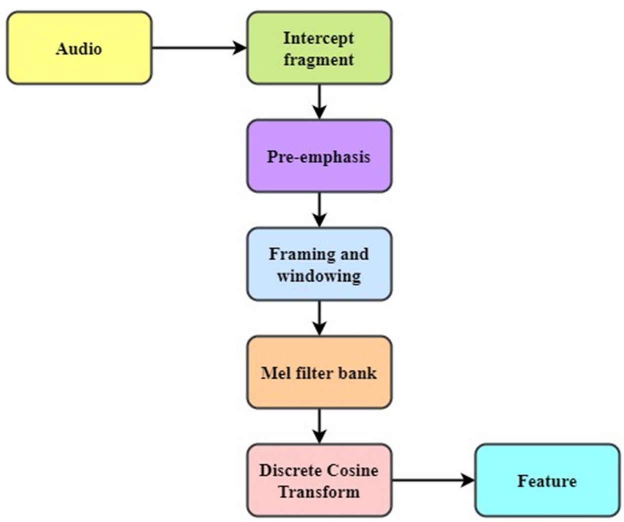


Fig. 2 The workflow of the proposed EMRCF framework

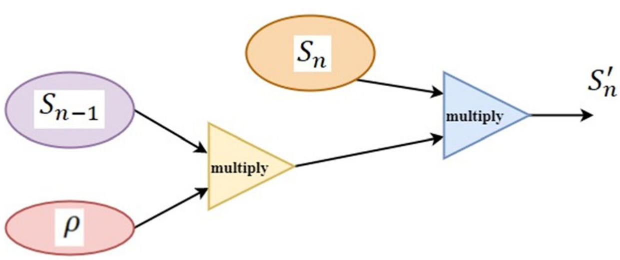


Fig. 3 Pictorial representation of S'_n

previously obtained variable is denoted as S_{n-1}. The focus factor is denoted as ρ. The predicted function for the next iteration is represented as S'_n. The high-frequency vibration maximum is illuminated. The pre-emphasis converter variable audio signals enhance the increased component to correct the significant proportion of the deleted audio signals.

(3) Analyzing and wavelet transform: the sound wave usually is continuously changing. The sound wave is thought to not shift significantly in one picture in a brief period to improve calculations. A Hanning window is applied to each transmission image to convert discreet Fourier to lessen the signal strength leaking.

(4) Mel filter function: The measure Meyer includes the individual hearing perceptual features. As reducing the size, the duration falls shorter, the rate is more significant, and the course longer. In general, the energy range is fitted with 10–30 rectangle detectors. Each buffer multiplies the amplification within each sample individually, and the calculated values are the intensity quality of defined planning in the machine's equivalent range of frequencies.

(5) Discrete Fourier transformations (DCT): consider a log transformation of the power results calculated using DCT to retrieve confessions on the lower frequencies explicitly, eliminate the similarity between the different diameters patterns and project a message into a reduced feature room. The Mel Frequency Coefficient function has been acquired.

3.2 Quantizing of vectors

The Features Extraction strategy is a software encoding and computing methodology that brings together various information and then measures them within the subspace, consolidating the capability to maintain no loss of detail. The suggested VQ layout algorithms, called the LBG-VQ application, are a learning sequence process. The m software matrix method creates:

(1) Pick m variables automatically and assign themselves as the original code variable from the character's amount.

(2) Grab the closest software variable by the Extracted features and assign it into the software vectors group, with each parameter in the specified function amount;

(3) Determine each nearest cluster variable and utilize it as a new search variable, reversing the two steps.

(4) If the cumulative deviation of both instances exceed the limit or the overall simulation maximum is reached, the code variable performance is halted. The function coefficients are separated into m groups after variable quantification, and the sound characteristic functions are determined. The P own quality comprises three components: the median variable per subclass μ_{P_i}, each kind of correlation coefficients $\sum P_{m_i}$ And the number of variables in each group as W_{P_i} is expressed in Eq. (2):

$$P = \left\{ \left(\mu_{P_1}, \sum P_1, W_{P_1} \right), \ldots, \left(\mu_{P_m}, \sum P_m, W_{P_m} \right) \right\} \tag{2}$$

The P own quality comprises three components: the median variable per subclass μ_{P_i}, each kind of correlation coefficients $\sum P_{m_i}$ the number of variables in each group as W_{P_i}. The range between variables is K-L by the K-mean segmentation, and the last musical characteristic M is received, as seen in Eq. (2), where m is the classified amount, where A_i is the security strategy of each group of its meaning P, and σ_i is the P covariance measure is expressed in Eq. (3)

$$M = mA_i\sigma_i \tag{3}$$

m is the classified amount, where A_i is the security strategy of each group of its meaning P, and σ_i is the P covariance measure.

3.3 Quantizing of vectors

The difference between the musicians is quantified using the Ground Mover's range equation after the function vectors repository is built. The less range, the stronger the resemblance. The EMD range is the minimal transmission expense under some restrictions and can reflect two multifaceted populations' similarities. The transmission distances between expected values are considered.

The music P function magnitude is P. This function is the student attribute in that music $P = \{(M_{P_1}, W_{P_1}), \ldots, (M_{P_m}, W_{P_m})\}$. The category mass is the category phenomenon, and the music Q is equivalent in the amount $= \left\{(M_{Q_1}, W_{Q_1}), \ldots, (M_{Q_m}, W_{Q_m})\right\}$. $D = d_{ij}$ is the surface range vector among M_{P_i} and M_{Q_i} and the base price feature for daily expenditure is as indicated in Eq. (4)

$$W = \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} f_{ij} \tag{4}$$

d_{ij} is denoted as the surface range vector among M_{P_i} and M_{Q_i} And the base price feature for expenditure. The recommendation function is denoted as f_{ij}. The weight factor of the neural network is denoted as W.

The following restrictions should be met. Equation (5) indicates that one path from P to Q can be relocated.

$$f_{ij} \geq 0 \tag{5}$$

The recommendation function is denoted as f_{ij}. It should be above zero value. Equations (6) and (7) show that the device must be less than its capacity and not more than its weight of the neural network should be the volume.

$$\sum_{i=0}^{m} f_{ij} \leq W_i \tag{6}$$

$$\sum_{j=0}^{m} f_{ij} \leq W_{Q_j} \tag{7}$$

The music P function magnitude is P. This function is the student attribute in that music $P = \{(M_{P_1}, W_{P_1}), \ldots, (M_{P_m}, W_{P_m})\}$. The category mass is the category phenomenon, and the music Q is equivalent in the amount $= \left\{(M_{Q_1}, W_{Q_1}), \ldots, (M_{Q_m}, W_{Q_m})\right\}$. The weight of the network is denoted as W_i and the weight of the link between P and Q is expressed as W_{Q_j}. Equation (8) equals the value of the weighted not exceeding the maximum usage of digital devices.

$$\sum_{i=1}^{m} \sum_{j=1}^{m} f_{ij} = \min\left(\sum_{i=1}^{m} W_{P_i}, \sum_{j=1}^{m} W_{Q_j} \right) \tag{8}$$

The recommendation function is denoted as f_{ij}. The weight of the network is denoted as W_i and the weight of the link between P and Q is expressed as W_{Q_j}. Normalization as in Eq. (9) after measuring the workflow F:

$$EMD(P, Q) = \frac{W(P, Q)}{\sum_{i=1}^{m} \sum_{j=1}^{m} f_{ij}} \tag{9}$$

The emotionally detected data between P and Q is expressed as $EMD(P, Q)$. The recommendation function is denoted as f_{ij}. The weight of the network is denoted as W_i and the weight of the link between P and Q is expressed as W_{Q_j}. The weight of the link between P and Q is expressed as $W(P, Q)$.

Figure 4 depicts the pictorial representation of $EMD(P, Q)$. The emotionally detected data between P and

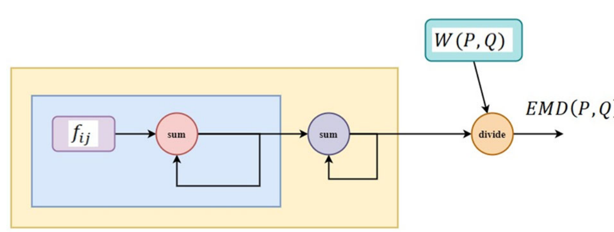


Fig. 4 Pictorial representation of $\boldsymbol{EMD(P, Q)}$

Q is expressed as $EMD(P,Q)$. The recommendation function is denoted as f_{ij}. The weight of the network is denoted as W_i and the weight of the link between P and Q is denoted as W_{Q_j}. The weight of the link between P and Q is expressed as $W(P,Q)$.

Researchers look at the use of speech feelings perception of convolutionary neural networks (CNNs). Networks are established to model the neural network in the visual analysis. Still, it is essential to optimize a channel for accurate calculations, considering the technical specifications and complexities. Consequently, a network is developed for a computer program that categorizes sentiment effectively in 4 feelings: joy, depression, aggressive and positive.

3.3.1 Overview of the database

The sample population included for the feature learning is a face recognition problem for Kaggle speech and the figures comprise 52 photographs of individuals with a luminance of 52 pixels. Each face is organized into eight courses of emotions: rage, resentment, terror, happiness, sorrow, disappointment and neutrality. Researchers used four feelings for these studies: angry, joy, sadness and detachment. Besides, 26,000 pictures match these feelings. The distribution is as obeys: 9000 specimens are pleased, 6000 representatives are unhappy, 6200 examples favorable, 5000 samples are irritated.

3.3.2 Characterization of prototype

A multi-layered convoluted computer program is configured to test customer image functionality. Source layer, convolutionary layer, hidden layer, output layer with weight matrix and feature extractor are used in the convolutionary computer program. This function is a continuous series of plates.

(1) Input layer: It has a defined and default size. Researchers employed OpenCV for facial object recognition before the picture was added to the surface to well before the image. During the Haar Streams, filtering along with Algorithm is used to identify and cultivate the mask quickly. The trimmed cover is then transformed into a 52-by-52-pixel gray. The proportions from RGB to gray can conveniently be loaded into the hidden layers as a numerical sequence. This phase is considerably reduced.

(2) Convolutionary layers: Another neural network in the convolution layer is several separate kernels or function detection systems with workouts produced by irregular intervals. Each function detection system is considered an external that slides through the actual picture and calculates a graph.

For the same visual information, the activation function produces various function charts. Different detectors have been used to execute procedures that show that blurred and feature extraction pixel is improved. Several input map data are progressively implemented over the whole picture.

Each convolutionary layer produces 512 function representations in the human brain. After each accumulation period, the gated recurrent unit was included. Using the common grouping approach Max Consolidation after a series of convolutionary layers, each function map's orbital angular momentum has been reduced while preserving essential details. Researchers utilized browsers that only take into account the full image pixels in the function graph view. The particles form a picture with lower losses by 4.

(3) Dense layers: The convolutionary and max-pooling contribution is a high-ranking picture element. These characteristics are used in the hidden layers to group the data into multiple groups. The features are translated using layers linked to the volumes.

The system is designed by the spread of information sets and the reversal of mistakes. The proposed emotion-based music recommendation and classification framework (EMRCF) uses three linked concurrent layers. The approach easily generalizes to new pictures and can adapt progressively to reduce waste. A 30% randomization was introduced to avoid the over-alignment of performance appraisals. It enabled us to monitor the system's exposure to loudness while retaining the design and architecture sophistication.

(4) Output Layer: Autoencoder is included in a deep network's hidden layers as the complex formation. The performance is then viewed for each emotional category as a normal distribution.

Figure 5 depicts the core architecture of the proposed EMRCF Framework. The Neural Networks framework was used to practice and simulate possible algorithms' permutations to use an 850 M visuals module. The learning curve and validity of the measurement are significantly diminished. The core network eventually comprises nine activation functions of residual blocks, supplemented by two large pieces after all three convolutionary layers.

3.3.3 Outcome

The last framework comprises 20,000 videos and 5000 photographs. The template ultimately reached the

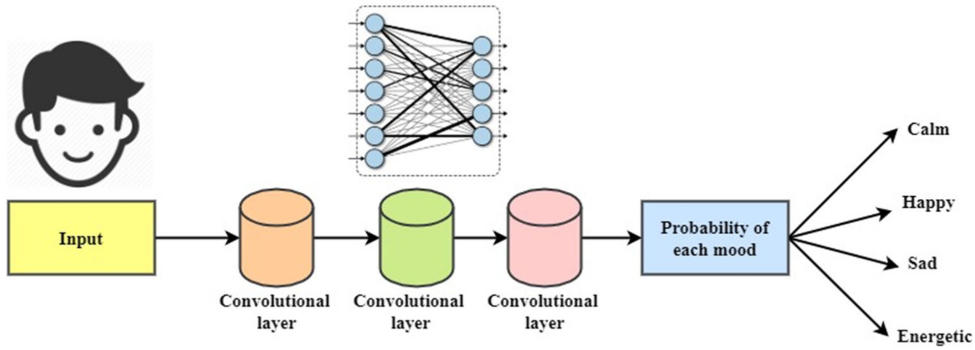


Fig. 5 The core architecture of the proposed EMRCF framework

discriminant function for the system with a precision of 91.23 percent. Obviously, in the classification of photographs in the type of "frustrated," the device does very efficiently. Researchers notice fascinating findings in the categories "good" and "depressing" due to the significant variations in feature vectors as described. The software's F-measurement is 92.12 percent.

3.4 Module for Music Identification

In this part, researchers explain how each music was mapped to its emotions. With audio pitch and other modern software processing techniques, researchers obtained the vibration parameters of the tracks. Focused on these characteristics, researchers have educated an adaptive neural system that categorizes the music with 96.12 percent accuracy in four sections succeeds.

Figure 6 shows the music recommendation system of the proposed EMRCF framework. Researchers explain how each music was mapped to its emotions. With audio pitch and other modern software processing techniques, researchers obtained the vibration parameters of the tracks. Focused on these characteristics, researchers have educated an adaptive neural system that categorizes the music with 96.12 percent accuracy in four sections succeeds.

3.4.1 Dataset description

The database contains 400 music in four ways. Category A with 120 music, Category B with 100 pieces, Category C with 80 music and Category D with 100 themes are all distributed. The tracks were marked automatically, and 15 paying topics were checked for the different classifiers. Category A includes thrilling and powerful music; Category B includes cheerful and joyous music; Category C includes dealcoholizing and unhappy music, while category D offers peaceful and comfortable pieces.

(1) Pre-processing: together, all compositions have been processed at a standardized frame rate of 256 kbps (250–120 Hz), with a single sound stream. Researchers break a little more each track to get clips containing the most significant sections of the way. The variables were generalized such that the median and component variability were null.

(2) Summary: After watching the news studies and the findings of the 2015 Audio Condition Identification challenge, researchers have found some texture numerous psychological. The abstraction method's applicant characteristics pertain to distinct classifications: harmonic, syncopated, melodic (Chords, Pulse continuum and so forth), harmonic and pitch (modes, RMSE and roll-off). These are just standard

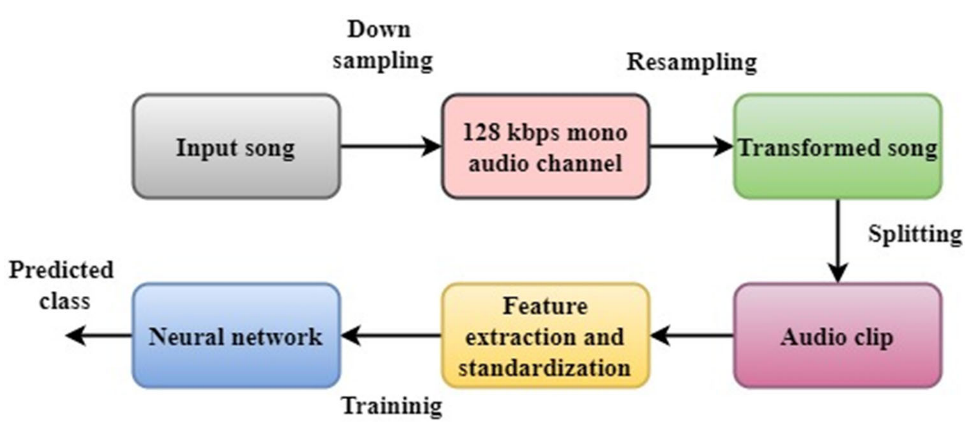


Fig. 6 The music identification system of the proposed EMRCF framework

definitions. The functions and the appropriate applications were collected with Python 5.

Researchers have used Repetitive Function Removal, after specifying all the settings, to pick the functionality that better improves the proposed system RFE's reliability by remedying characteristics and modeling those remaining characteristics. The template precision is used to determine the features and feature values most related to the goal parameter forecast. Sound, harmonic roll-off, cepstral correlations, speed, average root squared power, central procedural space, beat the range and lower layout, Fourier transformation and the music's kurtosis are all characteristics.

3.4.2 Overview of design

A multi-layer neuronal system assessed the music's attitude. The network has an entrance layer, some hidden layers and thick convolution layers. Set and default measurements are present in the hidden layers. It accepts the 20 variables as part of implementing the function to ensure that the database is not dimensional. The template was thus useful in actual situations.

The hidden layer is a conventional multi-layer sensor, which allows one to combine functionality with improving the precision of the identification. The hidden layers utilize a soft-max amplification feature to provide each emotion category's performance as a likelihood.

3.4.3 Results

After 15 times of cross-validation utilizing our computational model, researchers obtained a total precision of 94.69 percent and 96.69 percent of the F1 score. The vector for uncertainty. The song identification element has a too high production quality, no question.

3.5 Recommendation module

This framework produces a music track of the associated channels. It helps the operator change the music track depending on their choice and even change the track's training data.

Figure 7 shows the music recommendation system's workflow of the proposed EMRCF Framework. This framework produces a music track of the associated channels. It helps the operator change the music track depending on their choice and even change its training data.

3.5.1 Development of maps and playlists

Music categorized are identified to the emotions of the participant. An album of related music is created after the visualization process is done. During soundtrack generation, different songs are clustered collectively. Comparing music surrounded by white 100 ms, centered around each 15 ms timespan, determined the music's commonality. Researchers observed the length of such periods according to a standard music statement's intensity after observational analyses. Sound intensity, tempo, tone and smoothness are the factors associated with all music. This music helps to group the songs for the recommendation. The soundtrack is generated as melody, harmony and dancing music with a low, medium and high pitch based on these commonalities. Various combinations are identified and clustered for soundtrack generation.

The vector separation method was employed to evaluate the resemblance between music clips. Obtained features of a recording were contrasted with the same value with music clips of the same category mark with much the same functions. There is a double function in the search algorithm; its recommendations music dependent upon:

1. Powerful atmosphere for the consumer.
2. Choice of the consumer.

At first, a soundtrack is created of all the tracks of the given category. Based on your preference, the individual can choose a piece of music as a favorite. The meaningful matter in the song list is given to a favorite track. The tone of a song may be interpreted differently by individuals. The consumer will adjust the organizations' compositions as per his preference for music if he understands this.

3.5.2 Music app evolution

Using a standard online training method, the Extreme Gradient Descend, researchers were ready to enact an interactive media player (SGD). If the client needs to modify a music's classification, SGD considers the new mark for that consumer. A music's frequency or strength can be associated with rage, whereas gentle music can be associated with softness, sorrow or anxiety. A music's total spike can indicate fun, relaxed and pleasant emotions, while a music's optimal peak may indicate a gloomy, negative and painful sound.

Many single-pass results are tested with our method for accuracy, and the portable device's real-time design is the most combined with SGD. After analyzing some data point in the database, variable updation in SGD exists. This method gives the tracking number regression method of great benefits. Initially, it reduces the time taken to calculate massive data collections' prices and differential. Furthermore, it is better to integrate current information or

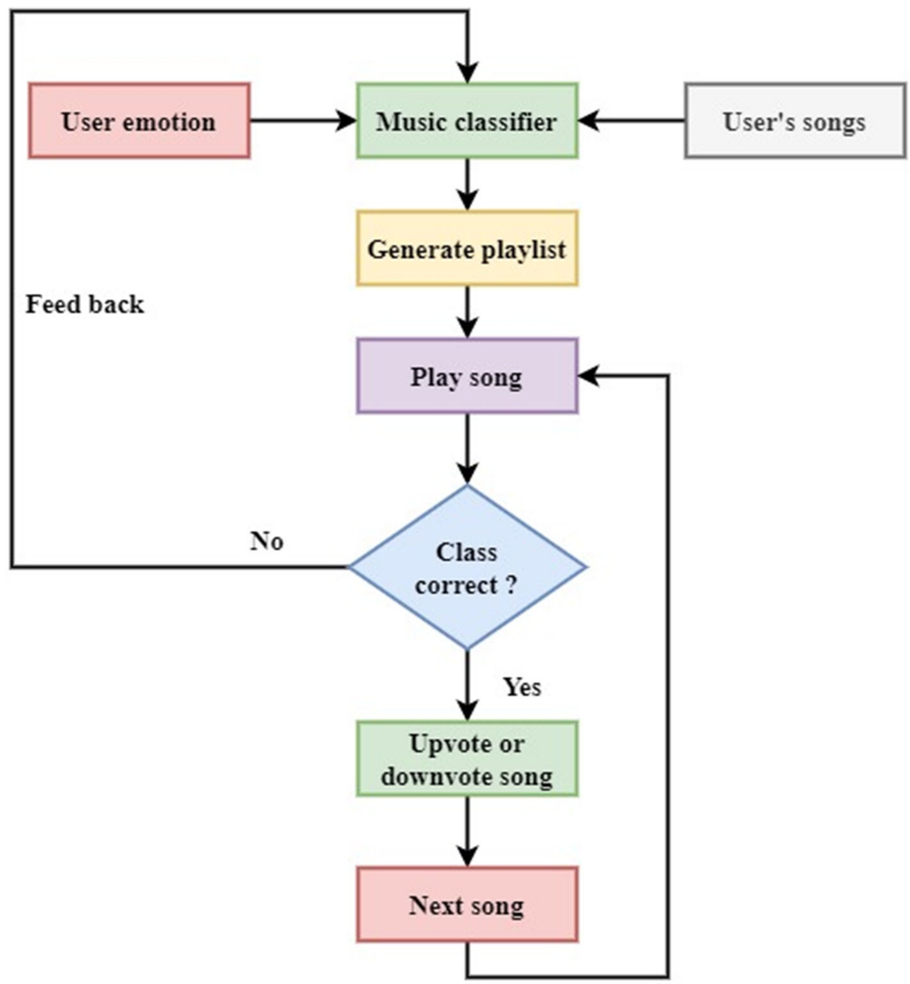


Fig. 7 The workflow of the music recommendation system of the proposed EMRCF framework

to update previous equipment. The periodic, variable changes require a lesser information gain than the batch loss function.

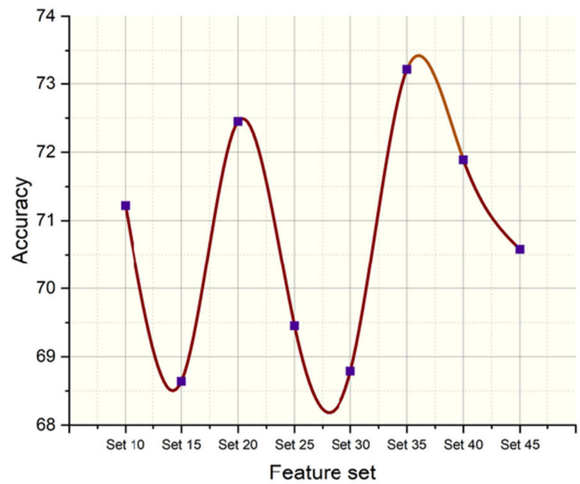
4 Software analysis and performance evaluation

The proposed emotion-based music recommendation and classification framework (EMRCF) is designed and analyzed. The different simulation parameters, such as accuracy, precision, efficiency, recommendation rate and recall rate, are explored. The results show that the proposed emotion-based music recommendation and classification framework (EMRCF) had the highest performance because of machine learning and Internet of things.

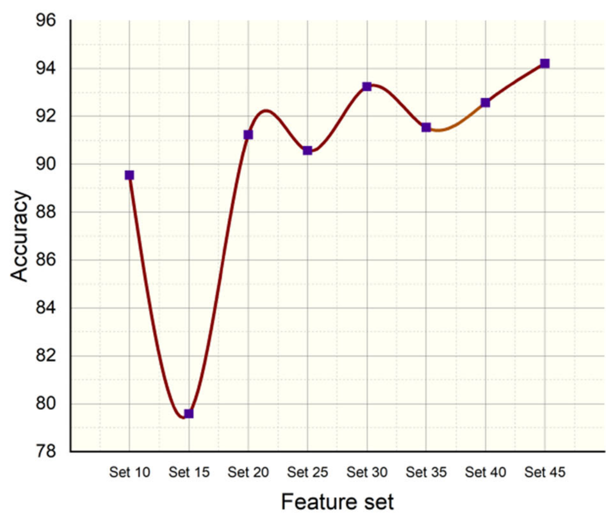
Figure 8(a and b) shows the existing Support Vector Machine's accuracy analysis and the proposed EMRCF framework, respectively. The dataset is varied from set 15 to ser 45 with a step size of 5 for the simulation analysis. The respective accuracy performance is calculated and measured for both the existing and the proposed EMRCF framework. The results show that the proposed EMRCF Framework) has the highest accuracy in all situations.

Table 1 shows the Mean Error Analysis of the proposed EMRCF framework. The dataset is varied from set 10 to set 45 with a step size of 5 for the simulation analysis. The advanced classification algorithms such as decision tree (DT), random forest (RF), support vector machine (SVM) and neural network (NN) are analyzed in the software. The data set is classified as frustrated, pleasant and relaxed. The music recommendation is predicted for the validated as a group. The results show that the neural network has the highest performance. The research incorporates the neural network with the proposed EMRCF framework.

Figure 9(a and b) shows the efficiency analysis of the existing PPG method and the proposed EMRCF framework, respectively. The different classification algorithms such as decision tree (DT), random forest (RF), support vector machine (SVM) and neural network (NN) are considered for the simulation analysis. The respective efficiency for the existing and the proposed system is



(a) **Accuracy analysis of the existing Support Vector Machine**

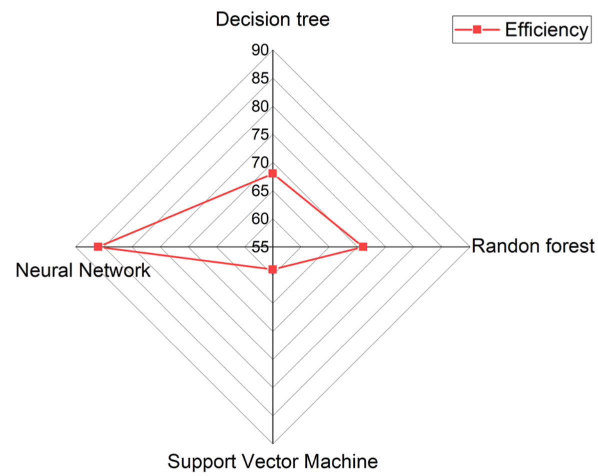


(b) **Accuracy analysis of the proposed EMRCF Framework**

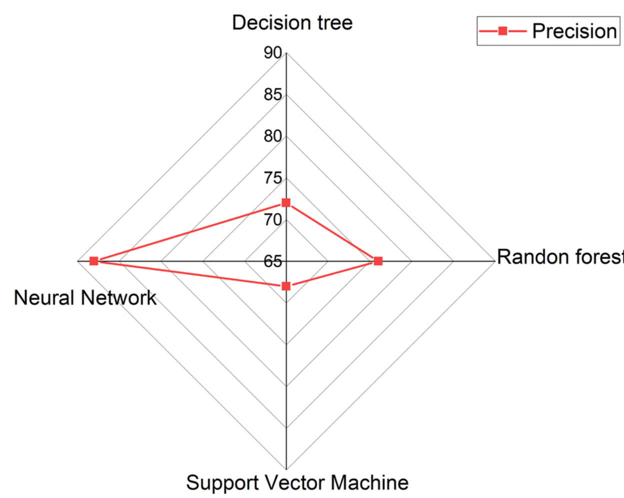
Fig. 8 (**a**) Accuracy analysis of the existing Support Vector Machine. (**b**) Accuracy analysis of the proposed EMRCF framework

Table 1 Mean average error analysis of the proposed EMRCF framework

Data set	DT	RF	SVM	NN
10	1.34	1.21	1.12	1.02
15	1.22	1.10	1.01	0.96
20	1.12	1.02	0.98	0.91
25	1.05	0.96	0.96	0.89
30	0.98	0.92	0.91	0.82
35	0.95	0.86	0.86	0.79
40	0.86	0.81	0.82	0.75
45	0.82	0.78	0.75	0.71



(a) **Efficiency analysis of the existing PPG method**



(b) **Efficiency analysis of the proposed EMRCF Framework**

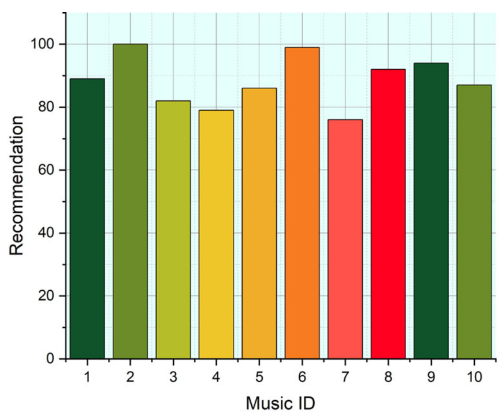
Fig. 9 (**a**) Efficiency analysis of the existing PPG method. (**b**) Efficiency analysis of the proposed EMRCF Framework

measured and plotted in the above figure. The results show that the proposed emotion-based music recommendation and classification framework (EMRCF) has the highest performance all the time.

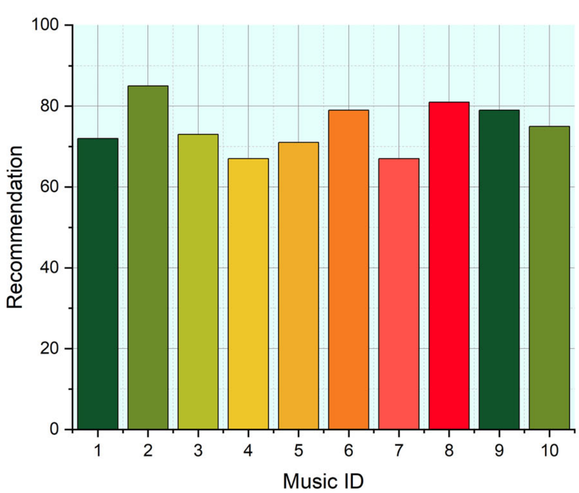
Table 2 shows the proposed EMRCF Framework. The window duration size for the simulation is varied from 5 to 40 with the step size of 5 windows. The different classification algorithms such as decision tree (DT), random forest (RF), support vector machine (SVM) and neural network (NN) are considered for the simulation analysis. The results show that the proposed EMRCF framework, a neural network, has the highest performance in all the situations to the existing methods.

Table 2 Recall rate analysis of the proposed EMRCF framework

Window duration size	DT	RF	SVM	NN
5	69.5	68.5	67.8	66.5
10	70.4	69.4	69.8	64.5
15	70.8	69.8	70.2	67.5
20	70.1	70.5	71.2	66.8
25	68.5	71.2	68.2	64.9
30	64.7	69.5	69.4	67.2
35	70.5	68.7	67.3	65.1
40	69.8	69.2	69.8	65.4



(a) Recommendation analysis of the proposed EMRCF Framework



(b) Recommendation analysis of the existing PPG system

Fig. 10 (**a**) Recommendation analysis of the proposed EMRCF Framework. (**b**) Recommendation analysis of the existing PPG system

Figure 10(a and b) shows the recommendation analysis of the proposed EMRCF framework and the existing PPG system, respectively. The list of songs in the music track is considered for the simulation analysis. Based on the participant's mood analysis, the song is recommended for the play. The song which has the highest recommendation ratio will be played next. The results show that the proposed EMRCF framework has the highest efficiency in all situations.

The proposed EMRCF framework is designed and analyzed. The different simulation parameters, such as accuracy, precision, efficiency, recommendation rate and recall rate, are explored. The results show that the proposed EMRCF framework has the highest performance because of machine learning and Internet of things.

5 Conclusion and future scope

The abovementioned findings are quite successful. The individual user's high precision and good reproducibility render it ideal with most real-world applications. The framework for the song identification is beneficial; it achieved high accuracy in the group "frustrated," and in the classes "pleasant" and "relaxed." However, the proposed emotion-based music recommendation and classification framework (EMRCF), consumer attempts to generate soundtracks are reduced. The space to develop is often recognized. It is important to assess why the program reacts by considering all seven real feelings; the suggestions framework may be strengthened with different music from foreign dialects and areas. A clustering algorithm can gather user tastes to enhance the whole scheme. Researchers focused the music recommendation based on past history and present mood in the proposed development through classification and clustering mechanisms. The EMRCF performs well with 96.12% of prediction accuracy. The good outcomes for the fourth humor studies are achieved by converting participant's emotions effectively to the right music classification with an average precision of 96.69 percent.

Acknowledgements The authors would like to acknowledge the support of Taif University Researchers Supporting Project number (TURSP-2020/292), Taif University, Taif, Saudi Arabia. The authors also would like to acknowledge the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the fast-track Research Funding Program.

Author contributions Conception and design of study: MTQ; Acquisition of data: EHA; Analysis and/or interpretation of data: MHK; Drafting the manuscript: MAK.

Funding This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program. This work was supported by Taif University Researchers Supporting Project number (TURSP-2020/292), Taif University, Taif, Saudi Arabia.

Declarations

Conflict of interest I have no conflicts of interest to report regarding the present study.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abdul A, Chen J, Liao HY, Chang SH (2018) An emotion-aware personalized music recommendation system using a convolutional neural networks approach. Appl Sci 8(7):1103

Andjelkovic I, Parra D, O'Donovan J (2019) Moodplay: Interactive music recommendation based on Artists' mood similarity. Int J Hum Comput Stud 121:142–159

Ayata D, Yaslan Y, Kamasak ME (2018) Emotion-based music recommendation system using wearable physiological sensors. IEEE Trans Consum Electron 64(2):196–203

Chen J, Ying P, Zou M (2019) Improving music recommendation by incorporating social influence. Multimed Tools Appl 78(3):2667–2687

Domínguez-Jiménez JA, Campo-Landines KC, Martínez-Santos JC, Delahoz EJ, Contreras-Ortiz SH (2020) A machine learning model for emotion recognition from physiological signals. Biomed Sign Process Control 55:101646

Er MB, Aydilek IB (2019) Music emotion recognition by using chroma spectrogram and in-depth visual features. Int J Comput Intell Syst 12(2):1622–1634

Guo L, Liang J, Zhu Y, Luo Y, Sun L, Zheng X (2019) Collaborative filtering recommendation based on trust and Emotion. J Intell Inform Syst 53(1):113–135

Hsu JL, Zhen YL, Lin TC, Chiu YS (2018) Affective content analysis of music emotion through EEG. Multimedia Syst 24(2):195–210

Kim B, Lee J (2018) A deep-learning-based model for emotional evaluation of video clips. Int J Fuzzy Log Intell Syst 18(4):245–253

Kim HY, Lim SB (2018) Emotion-based Hangul font recommendation system using crowdsourcing. Cogn Syst Res 47:214–225

Krämer B (2018) Online music recommendation platforms as representations of ontologies of musical taste. Communications 43(2):259–281

Li C, Bao Z, Li L, Zhao Z (2020) Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. Inf Process Manage 57(3):102185

Lu F, Tintarev N (2018) A diversity adjusting strategy with personality for music recommendation. In: IntRS@ RecSys (pp 7–14)

Murthy Y, Koolagudi SG (2018) Content-based music information retrieval (CB-mir) and its applications toward the music industry: a review. ACM Comput Surv (CSUR) 51(3):1–46

Rachman FH, Sarno R, Fatichah C (2018) Music Emotion classification based on lyrics-audio using corpus-based emotion. Int J Electric Comput Eng 8(3):2088–8708

Schedl M, Zamani H, Chen CW, Deldjoo Y, Elahi M (2018) Current challenges and visions in music recommender systems research. Int J Multimed Inform Retriev 7(2):95–116

Seo YS, Huh JH (2019) Automatic emotion-based music classification for supporting intelligent IoT applications. Electronics 8(2):164

Volokhin, S., & Agichtein, E. (2018, March). Understanding music listening intents during daily activities with implications for contextual music recommendation. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 313–316).

Wang D, Deng S, Xu G (2018a) Sequence-based context-aware music recommendation. Inform Retriev J 21(2):230–252

Wang D, Deng S, Zhang X, Xu G (2018b) Learning to embed music and metadata for context-aware music recommendation. World Wide Web 21(5):1399–1423

Yang X, Dong Y, Li J (2018) Review of data features-based music emotion recognition methods. Multimedia Syst 24(4):365–389

Zangerle E, Chen CM, Tsai MF, Yang YH (2018) Leveraging effective hashtags for ranking music recommendations. IEEE Trans Affect Comput