Novel mathematical model for the classification of music and rhythmic genre using deep neural network

Swati A. Patil[1,2*], G. Pradeepini[1] and Thirupathi Rao Komati[3]

*Correspondence:
swati.patil251@gmail.com

[1] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, A.P. 522502, India
[2] G. H. Raisoni Institute of Engineering and Business Management, Jalgaon 425001, India
[3] GST GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh 530045, India

Abstract

Music Genre Classification (MGC) is a crucial undertaking that categorizes Music Genre (MG) based on auditory information. MGC is commonly employed in the retrieval of music information. The three main stages of the proposed system are data readiness, feature mining, and categorization. To categorize MG, a new neural network was deployed. The proposed system uses features from spectrographs derived from short clips of songs as inputs to a projected scheme building to categorize songs into an appropriate MG. Extensive experiment on the GTZAN dataset, Indian Music Genre(IMG) dataset, Hindustan Music Rhythm (HMR) and Tabala Dataset show that the proposed strategy is more effective than existing methods. Indian rhythms were used to test the proposed system design. The proposed system design was compared with other existing algorithms based on time and space complexity.

Keywords: Mathematical model, Neural network, Feature extraction, MG, Time complexity, Space complexity, Deep neural network Learning, GTZAN

Introduction

Currently, huge amounts of data are being produced on many community media platforms. Because the amount of data is so large in terms of bytes, it is critical to categorize it so that users can simply search for and promote it. Our work is mostly concerned with the song data. Music-like audio can be processed or evaluated as a song melody using instruments only, only acoustic songs without electric instruments, or songs with both music and instruments. Audio or musical gestures must be converted to a binary format before being processed on a computer. Audio data is an analog signal, which is converted to a digital format by an analog-to-digital converter. The spectrogram and waves depict the audio signals.

Currently, music or tunes are broadly scattered on the Internet, and the measures of transferring or discharging music/melodies are simply expanding and not halting or reducing. Effective association and processing of such huge amounts of information, just as effective ordering, looking, and recovery of such enormous measures of musical data is a major challenge. Most music arrangement techniques are based on music classified

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 2 of 17

by classification, state of mind, or various semantic tags [1]. The music Genre is a significant angle used to classify music/melodies in music shops, broadcasting, and searching. Automatic music sort characterization is a crucial part of music data recovery systems [2]. For humankind, it is not hard to order music in various types,however, comprehending it using PCs is a major test. There is a distinctive view that characterizes instruments, songs, beats, and moods to separate musical genres [3].

Music Genre grouping is generally significant in the music database of the executive's framework to look through music according to class. The kind order is a profoundly free type of naming [4]. Because of the specialized turn of events and innovations of savvy compact gadgets such as cell phones, pod music has reached everyone [5]. We as often as possible run over individuals with earphones or ear plugs in transit while strolling, driving even while examining, while at the same time doing activity or yoga. As the Internet is accessible, music has been effectively and promptly opened to individuals. When we listen to melodies, we never need to listen to the full playlist. Commonly, we need to listen to melodies that suit our temperament. Therefore, in this situation, it is significant or advantageous to have a rundown of melodies according to temperament or sort. Si that the client can pursue the music from the immense database without much of a stretch.

Liu et al. [6] have worked in the categorization of MG based on visual representation and have been effectively studied in recent years. Convolutional neural networks (CNN) have recently gained popularity as a means to accomplish tasks. The proposed CNN architecture uses multi-scale time–frequency information, transferring more appropriate semantic elements to the decision-making layer to distinguish the genre of an unfamiliar music trunk. The trials are assessed using the Ballroom, GTZAN and Extended Ballroom benchmark datasets. The proposed approach can obtain 93.9 percent, 96.7 percent, and 97.2%, respectively, according to the testing findings [7, 8].

The linking of genre-related labels to digital music files has been studied by [9] in MGC. It can be used to organize both commercial and private melody collections. Music tracks are frequently characterized as a collection of sound textures stimulated by timbres. To depict a complete track, the selection of sound textures is frequently used. They assessed the influence of texture selection on automatic MGC and provided a unique texture selector based on K-Means, which identifies a variety of sound textures inside each recording [10, 11].

Music is an essential and inseparable aspect of most people's lives. Several MG exist, each of which is distinct from the others, resulting in a wide range of musical preferences. Consequently, classifying music and recommending new music in music listening apps and platforms is an essential and timely subject. One of the most effective approaches to solving this challenge is to categorize music by genre. Several methods are available for categorizing and recommending music [12].

According to [13], music genre categorization is one of the sub-disciplines of music information retrieval (MIR) that is gaining in favor among academics, owing to the open problems. Despite the fact that research has produced a large number of publications, the issue still has a fundamental flaw: the music type is not clearly defined. Music classifications are hazy and ambiguous, owing to human subjectivity and disagreement [14].

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 3 of 17

Using a transfer learning technique [15], proposed an audio-based categorization strategy for 11 western MGs, including rock, pop, rape, country, metal, jazz, blue, R&B, folk music, Electronic Music, and Classical Music. One can achieve 0.9799 ROC-AUC and 0.8938 PR-AUC on a private 1100 songs collection. Using a deep learning model [16], created an autonomous music genre categorization. CNN is used for local feature extraction, whereas Long Short-Term Memory (LSTM) sequence-to-sequence autoencoders are used to learn representations of time series data while accounting for their temporal dynamics. Using the GTZAN dataset, computational experiments yielded an overall test accuracy of 95.4 percent and precision of 91.87 percent.

Labeling a music recording according to its genre is a natural and familiar method to characterize its content, according to [17] MG are useful for organizing music, creating individualized listening experiences, and creating playlists. Due to the inherent ambiguity and subjectivity, the automatic identification of an MG is a difficult task. Most music genre categorization efforts assume total label independence. Furthermore, the local classifier per node technique, which would be the best fit in this case, is time- and memory-intensive [18].

MGC is a vital and useful aspect of music information retrieval. Deep learning is increasingly used to classify MG for two reasons [19]. The starters eliminate the need for manual audio signal feature selection. Second, the hierarchical architecture is compatible with the temporal- and frequency-domain layering structure of music. The Transformer classifier performs better in MGC [20, 21] by analyzing the connection between distinct audio frames.

Jaime Ramirez [22] developed a web program that obtains tunes from YouTube and categorizes these tunes into MG. The technique described in this paper is built on models developed on audio sets' musical collection data. The conception output displays this sequential data in real time, in rhythm with the harmony video played the same time, with cataloguing outcomes displayed in the weighted area graph, where scores for the top-10 labels acquired for each block are given.

Automatic MGC is believed to be a basic component for obtaining music information, according to [22]. Indeed, music genre labels are extremely useful for categorizing albums, songs, and performers into groups with comparable features. To improve the recovery of MG, an accurate and effective MGC system is necessary. This research uses two key processes: feature extraction and classification, to present a novel music genre categorization model. Features such as "non-negative matrix factorization (NMF) features, Short-Time Fourier Transform (STFT) features, and pitch features were extracted during the feature extraction phase. Goulart et al. [23] proposed features extraction based on entropy fractal features and SVM classifier to classify blues, classical, and lounge genre achieved almost 100% accuracy.

Researchers Lima et al. [24] have proposed using genre labels to organize music, albums, and performers into groupings with common similarities. They describe a unique method for automatically determining the musical genre in Brazil based solely on the words of the songs. This type of categorization remains difficult in the difficulty.

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 4 of 17

To examine the influence of facet choice on the correctness of music genre categorization, Moses [25] utilized a Support Vector Machine with a radial basis function kernel as a classifier. In various feature combinations (FC), some traits are merged and matched (FC1, FC2, FC3, and FC4). The classification results demonstrated that the accuracy of each feature combination (FC) varied significantly, with the highest accuracy being 80 percent and the lowest being 67 percent. When the combination of the FC1 and FC2 features produces the same accuracy of 80%, the FC2 combination is preferred because the processing time is logically lower with fewer characteristics. For categorization, it is necessary to gather a large number of songs and index them according to their genre. The skip grammar model then locates comparable context songs for suggestions. Consequently, the proposed system functions as a complete music recommendation system that provides excellent user experience [26].

Lima et al. [27] Genre labels may be used to categorize songs, albums, and performers together based on their shared characteristics. In NLP, this type of categorization remains difficult. The F1-score parameter provides an average result of 0.48, and the BLSTM technique surpasses the other models in the studies. Gospel, funk-carioca, and sertanejo, which had F1 scores of 0.89, 0.70, and 0.69, respectively, may be classified as the most unique and easiest to categorize in the musical genres of Brazilian music settings [28].

They employed textual material given as lyrics of the accompanying song because genre dependence is not confined to the audio profile. To include acoustic information in spectrograms, they used a CNN-based feature extractor for spectrograms and Hierarchical Attention Network-based feature extractor for lyrics [29–31]. In several papers [32, 33], the active transfer MGC technique (ATMGCM) was proposed for musical genre categorization.

To create mathematical models for music genre categorization, [34] employed a machine learning technique created on a hybridizing genetic programming and neural networks. They created three multi-label classifiers with varied complexity and accuracy trade-offs that can detect the degree to which audio content belongs to 10 distinct MG.

According to [35], one of the goals of the music information retrieval (MIR) community is to study novel approaches and build innovative systems that can competently and successfully extract and propose songs from massive records of music content. They also show how our database can help with specific MIR tasks, and compare it to other databases available in the literature to help the MIR community.

According to the literature, genre classification requires extra attention to increase its performance metrics. The classification was based on data from the spectrograms of songs. The proposed approach will be incredibly useful in a number of industries, such as films and education. Numerous online music-playing services, such as Ganna. com, use MGC to improve user suggestions. Most of the research work is based on feature extraction techniques and then applying machine learning algorithms to classify the music or genres. In this work, because music is very dynamic in nature, manual extraction of features from music clips or by onset pattern detection or the music pattern at periodic time intervals does not provide accurate results. Therefore, a deep

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 5 of 17

neural network is used for the automatic detection of features that train as per the given musical pattern. Guido [36, 37] proposed a mass of discrete time signal based on Teager energy operator enables more precise and detailed analysis of audio signal. This author [36, 37] also proposed a model based on entropy and deep neural network to fuse the important features information. Scalvenzi et al. [38] proposed a method using discrete wavelet packet transform with support vector machine (SVM) to gain the accuracy near to 90% for sorting semantic information.

The remainder of this paper is organized as follows. "Proposed system architecture and mathematical model" section delves deeper into the mathematical model of the planned system. The recommended technique has been applied to a variety of musical styles and yielded positive results. The proposed system design is compared to a number of algorithms, and the results are reported in the Results and Discussion section. The proposed scheme is evaluated on multiple workstations to find out the time complexity of the proposed scheme, according to [39]. "Conclusion" section is conclusion part.
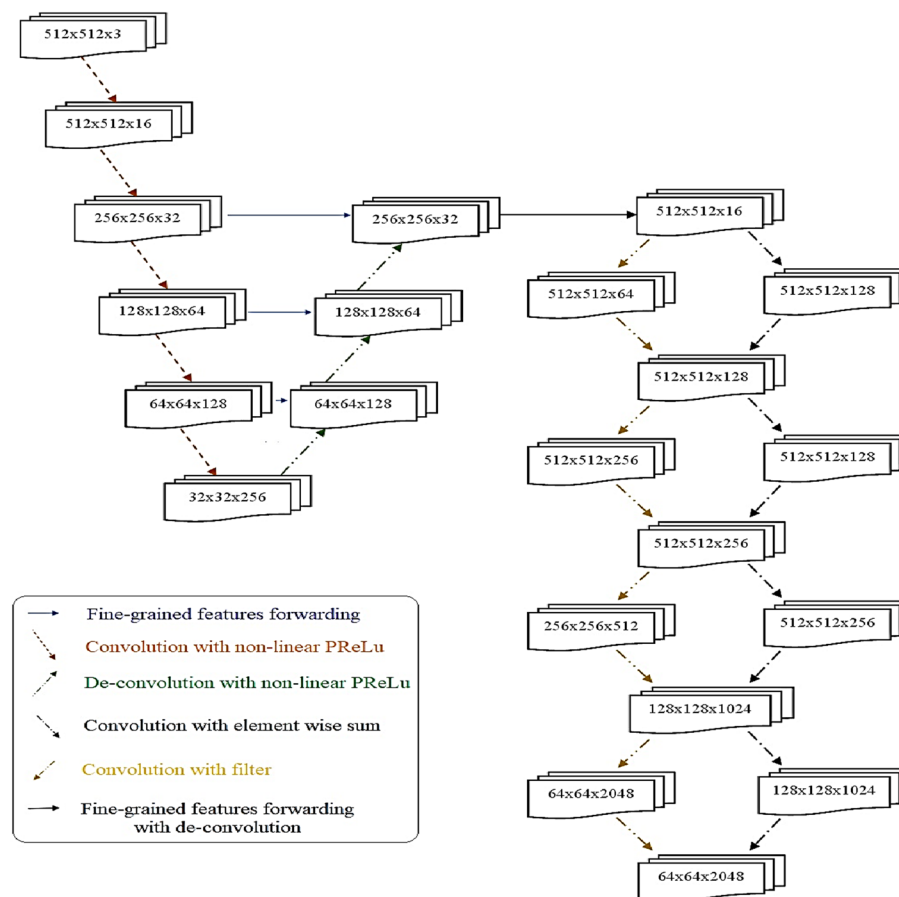
Fig. 1 Proposed system architecture of the MGC

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 6 of 17

Proposed system architecture and mathematical model

Figure 1 is the suggested system architecture. There are three components to the recommended architectural technique: compress, decompress, and filter CNN. Appropriate padding, with the purpose of both leveraging input features and decreasing resolution by using an appropriate stride at the end of each step run with convolutions. Fine-grained features forwarding provides the ideal estimate optimal fixed-lag. Optimal smoother provides the optimal estimate of x ˆ_t for a given fixed-lag N.

The main contributions of this research are as follows:

1. Extracting features using a Pre-trained Deep Neural Network (DNN) to evaluate the proposed method using precision, recall, F1-Score, and accuracy, and evaluating the system on selected datasets.
2. The parallel Convolution Deconvolution neural network model to extract finest level of features.
3. Removes the gradient descent problem using PreLU.

Steps in Proposed System Architecture:

The network architecture is decided on trial-and-error basis. The N no. of trials are carried out to observe the statical output. The hyperparameters: Number of input layer=1, layer 1=512, layer 2=256, layer3=128, layer4=64 neurons, number of output layer=1 and optimizer=adam.

1. In the first stage, convolution and deconvolution are performed. In deconvolution more features that are not important in convolution are extracted along with the features in the convolution. This mixture of the important and unimportant features is then fed to the convolution with filters and convolution with an element-wise sum. The proposed framework is illustrated in Fig. 1.
2. The reduction, decompression, and filtered CNN were the best methods.
3. All convolutions were performed with the required padding and, at the end of each step, a good stride to maximize the input features and reduce the resolution.
4. After additional convolution and pooling, the feature mapping decreased. Deconvolution is performed using the proposed methodology to obtain the original image, as shown in Fig. 1.
5. Deconvolution is the up-sampling of images. It generates high-resolution images and maps low-dimensional images to high-dimensional ones. A Parametric Rectified Linear Unit (PReLU) activation function is used in the convolution. PReLU handles negative inputs that the sigmoid activation function cannot handle. PReLU is a non linear activation function used as an activation function to make more differences among classes so that it improves the performance result.
6. The vanishing gradient problem was eliminated using PReLU.

The impulse response is obtained by multiplying the gaussian function with the sinusoidal wave. Due to multiplication-convolution condition, convolution of the Fourier Transform of the harmonic function and the FT of the Gaussian function (Convolution theorem) is the result of the Fourier transform (FT) of a filter's impulse response. In the filter, an actual and

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 7 of 17

an unreal component denote orthogonal orientations. The two pieces can be used together or independently to create a complicated number.

$$x(m,n; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{m'^2 + \gamma^2 n'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{m'}{\lambda} + \psi\right)\right) \tag{1}$$

$$x(m,n; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{m'^2 + \gamma^2 n'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{m'}{\lambda} + \psi\right) \tag{2}$$

$$x(m,n; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{m'^2 + \gamma^2 n'^2}{2\sigma^2}\right) \sin\left(2\pi \frac{m'}{\lambda} + \psi\right) \tag{3}$$

where Eq. 2.1, 2.2 and 2.3 is complex, real and imaginary respectively.

$$m' = m\cos\theta + n\sin\theta \text{ and } n' = -m\sin\theta + n\cos\theta \tag{4}$$

The wavelength of the sinusoidal component is represented by λ and the orientation of the normal to the parallel bands of a function is represented by θ in this equation. Here, σ is the sigma/standard deviation of the Gaussian envelope, ψ is the phase offset, and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the function.

$$\begin{bmatrix} \widehat{\mathbf{f}}_{t|t} \\ \widehat{\mathbf{f}}_{t-1|t} \\ \vdots \\ \widehat{\mathbf{f}}_{t-N+1|t} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \widehat{\mathbf{f}}_{t|t-1} + \begin{bmatrix} 0 & \cdots & 0 \\ \mathbf{I} & 0 & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{f}}_{t-1|t-1} \\ \widehat{\mathbf{f}}_{t-2|t-1} \\ \vdots \\ \widehat{\mathbf{f}}_{t-N+1|t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{U}^{(0)} \\ \mathbf{U}^{(1)} \\ \vdots \\ \mathbf{U}^{(N-1)} \end{bmatrix} g_{t|t-1} \tag{5}$$

where: $\widehat{\mathbf{f}}_{t|t-1}$ is assessed by filter; $\mathbf{g}_{t|t-1} = \mathbf{z}_t - \mathbf{H}\widehat{\mathbf{f}}_{t|t-1}$ is the novelty formed as the evaluation of the standard filter; the numerous $\widehat{\mathbf{f}}_{t-i|t}$ with $i = 1, \ldots, N-1$ are new variables; the improvements are calculated by the following scheme:

$$\mathbf{U}^{(i+1)} = \mathbf{Q}^{(i)} \mathbf{D}^{\mathrm{T}} \left[\mathbf{DQD}^{\mathrm{T}} + \mathbf{R}\right]^{-1} \tag{6}$$

and

$$\mathbf{Q}^{(i)} = \mathbf{Q} \left[(\mathbf{F} - \mathbf{UD})^{\mathrm{T}}\right]^i \tag{7}$$

where Q and U are the correlation of prediction error and the filter gains. The estimation error covariance is calculated in such a way that

$$\mathbf{Q}_i := E\left[\left(\mathbf{f}_{-i} - \widehat{\mathbf{f}}_{-i|t}\right)^* \left(\mathbf{f}_{-i} - \widehat{\mathbf{f}}_{-i|t}\right) \mid Z_1 \ldots Z_t\right] \tag{8}$$

then we get that the upgrading on the assessment of $\mathbf{f}_{t-i}$ is given by:

$$\mathbf{Q} - \mathbf{Q}_i = \sum_{j=0}^{i} \left[\mathbf{Q}^{(j)} \mathbf{D}^{\mathrm{T}} \left(\mathbf{DQD}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \mathbf{D}\left(\mathbf{Q}^{(i)}\right)^{\mathrm{T}}\right] \tag{9}$$

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 8 of 17

A feedback pass is also used, which analyses data reserved from the filter forward pass. The backward pass equations entail recursive data computations that are used to determine the smoothed state and covariance at each observation period. The recursive equations are given as,

$$
\begin{aligned}
\tilde{\Lambda}_k &= \mathbf{D}_k^{\mathrm{T}} \mathbf{S}_k^{-1} \mathbf{D}_k + \widehat{\mathbf{C}}_k^{\mathrm{T}} \tilde{\Lambda}_k \widehat{\mathbf{C}}_k \\
\widehat{\Lambda}_{k-1} &= \mathbf{F}_k^{\mathrm{T}} \tilde{\Lambda}_k \mathbf{F}_k \\
\widehat{\Lambda}_n &= 0 \\
\tilde{\lambda}_k &= -\mathbf{D}_k^{\mathrm{T}} \mathbf{S}_k^{-1} \mathbf{y}_k + \widehat{\mathbf{C}}_k^{\mathrm{T}} \widehat{\lambda}_k \\
\widehat{\lambda}_{k-1} &= \mathbf{F}_k^{\mathrm{T}} \tilde{\lambda}_k \\
\widehat{\lambda}_n &= 0
\end{aligned}
\tag{10}
$$

where $\mathbf{S}_k$ is the residual covariance and $\widehat{\mathbf{C}}_k = \mathbf{I} - \mathbf{U}_k \mathbf{D}_k$. Substituting the smoothed state and covariance into the equations yields the smoothed state and covariance.

$$
\begin{aligned}
\mathbf{Q}_{k|n} &= \mathbf{Q}_{k|k-1} - \mathbf{Q}_{k|k-1} \tilde{\Lambda}_k \mathbf{Q}_{k|k-1} \\
\mathbf{f}_{k|n} &= \mathbf{f}_{k|k-1} - \mathbf{Q}_{k|k-1} \tilde{\lambda}_k
\end{aligned}
\tag{11}
$$

For a random vector $\mathbf{f} = (f_1, \ldots, f_L)$, sigma points are any set of vectors

$$
\{\mathbf{S}_0, \ldots, \mathbf{S}_N\} = \{(S_{0,1}\, S_{0,2} \cdots S_{0,L}), \ldots, (S_{N,1}\, S_{N,2} \cdots S_{N,L})\}
\tag{12}
$$

attributed with first-order weights $W_0^a, \ldots, W_N^a$ that fulfill

$$
\sum_{j=0}^{N} W_j^a = 1
\tag{13}
$$

for all $i = 1, \ldots, L : E[x_i] = \sum_{j=0}^{N} W_j^a s_{j,i}$

Second-order weights $W_0^c, \ldots, W_N^c$ that fulfill

$$
\sum_{j=0}^{N} W_j^c = 1
\tag{14}
$$

for entire duos $(i, l) \in \{1, \ldots, L\}^2 : E[x_i x_l] = \sum_{j=0}^{N} W_j^c s_{j,i} s_{j,l}$.

A naïve choice of sigma points and masses for $\mathbf{x}_{k-1|k-1}$ in the filter algorithm is

$$
\begin{aligned}
\mathbf{S}_0 &= \widehat{\mathbf{f}}_{k-1|k-1} \\
-1 &< W_0^a = W_0^c < 1 \\
\mathbf{S}_j &= \widehat{\mathbf{f}}_{k-1|k-1} + \sqrt{\frac{L}{1 - W_0}} \mathbf{A}_j, \, j = 1, \ldots, L \\
\mathbf{S}_{L+j} &= \widehat{\mathbf{f}}_{k-1|k-1} - \sqrt{\frac{L}{1 - W_0}} \mathbf{A}_j, \, j = 1, \ldots, L \\
W_j^a &= W_j^c = \frac{1 - W_0}{2L}, \, j = 1, \ldots, 2L
\end{aligned}
\tag{15}
$$

Given prediction estimates $\widehat{\mathbf{f}}_{k|k-1}$ and $\mathbf{Q}_{k|k-1}$, a novel set of $N = 2L + 1$ sigma points $\mathbf{s}_0, \ldots, \mathbf{s}_{2L}$ with corresponding first-order weights $W_0^a, \ldots W_{2L}^a$ and second-order weights $W_0^c, \ldots, W_{2L}^c$ is computed. These sigma points are converted over h.

$$\mathbf{Z}_j = h(\mathbf{S}_j),\ j = 0, 1, \ldots, 2L \tag{16}$$

Then the experimental mean and covariance of the changed points are calculated.

$$\widehat{\mathbf{Z}} = \sum_{j=0}^{2L} W_j^a \, \mathbf{Z}_j$$
$$\widehat{\mathbf{S}}_k = \sum_{j=0}^{2L} W_j^c \left(\mathbf{Z}_j - \widehat{\mathbf{Z}}\right) \left(\mathbf{Z}_j - \widehat{\mathbf{Z}}\right)^{\mathrm{T}} + \mathbf{R}_k \tag{17}$$

where $\mathbf{R}_k$ is the covariance matrix of the observer features, $\mathbf{v}_k$. Moreover, the cross-covariance matrix is also wanted

$$\mathbf{C_{SZ}} = \sum_{j=0}^{2L} W_j^c \left(\mathbf{S}_j - \widehat{\mathbf{f}}_{k|k-1}\right) \left(\mathbf{Z}_j - \widehat{\mathbf{Z}}\right)^{\mathrm{T}} \tag{18}$$

where $\mathbf{s}_j$ are the unchanged sigma points created from $\widehat{\mathbf{f}}_{k|k-1}$ and $\mathbf{Q}_{k|k-1}$. The filter gain is given as,

$$\mathbf{U}_k = \mathbf{C_{SZ}} \, \widehat{\mathbf{S}}_k^{-1} \tag{19}$$

The updated covariance and mean estimates are

$$\widehat{\mathbf{f}}_{k|k} = \widehat{\mathbf{f}}_{k|k-1} + \mathbf{U}_k \left(\mathbf{Z}_k - \widehat{\mathbf{Z}}\right)$$
$$\mathbf{Q}_{k|k} = \mathbf{Q}_{k|k-1} - \mathbf{U}_k \widehat{\mathbf{S}}_k \mathbf{K}_k^{\mathrm{T}} \tag{20}$$

In the right-side part of the network's convolutional stages, we used stacking method where output of a layer is added to deep layer. The extracted features were then sent from the initial stages to final stage scanning from the left side of the system to the right segment through horizontal linkages, as shown in Fig. 1. So, we able to improve the precision of the final silhouette prediction by finding very fine data that would have been lost during the compression stage. The merging time of model has improved as a effect of these relations.

Data preparation

For implementation the proposed method is applied on two datasets. GTZAN, Hindustani Music Rhythm (HMR), and Indian Music Genre (IMG) datasets. GTZAN has 10 genres viz blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock have 1000 songs. HMR has 3 genres of Hindustani Music Rhythm dataset have taals: Ektaal, Zaptaal, Rupak, Teentaal have 300 songs and IMG dataset has 5 genres Sufi, bollypop, classical, Ghazals and Carnatic music have 500 songs. For feature extraction Librosa library is used. Keras, numpy, pandas library of python is used to implement proposed method. Training and testing data is selected in 80:20 ratio that is 80 as training and 20 as testing. For tonality perception using one of percussion instruments Tabala is used here as a dataset. This dataset contains total 561 files played on Tabala with Addhatritaal, bhajani, deepchandi, dadra, rupak, zaptaal, ektaal, and tritaal with 78,72,60, 72, 60, 60, 60,99 clips. The time signature of all

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 10 of 17
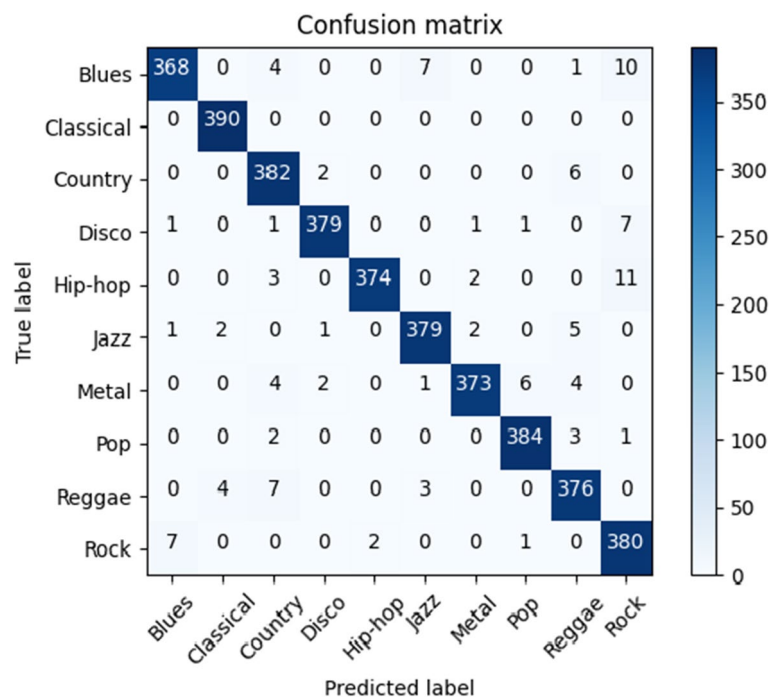
Fig. 2 The suggested algorithm's confusion matrix on the GTZAN dataset

Table 1 Performance parameters of proposed algorithm on GTZAN dataset

	Accuracy	Precision (%)	Recall (%)	F1-score (%)
Proposed system	99.41%	96.90	97.10	97.10
BiLSTM [43]	97.80%	88	94	78
4 Layers-2D CNN with Data Augmentation [44]	81.55	83.50	89.01	86.17
RCNN with Data Augmentation [44]	69.49	97.94	88.04	92.7

the taal in suggested method have different number of beats. Like Teentaal have 16 beats, Rupak have 7 beats, Ektaal 12 beats, Zaptaal have 10 beats.

Results and discussion

Planned approach is verified on 390 songs from very well-known dataset GTZAN, one from every category.

Figure 2 shows typical confusion matrix diagonal clearly contains dominating values, as shown by the confusion matrix. The confusion matrix also illustrates that the suggested design reduces the chances of misdiagnosis.

The performance characteristics have been determined using the confusion matrix. Accuracy, recall, and F1-score were calculated using standard mathematical techniques and presented in Table 1.

We can simply state that the recommended construction for the sorting of songs genre is dependable since the accuracy of the categorization of music genre ranges

Patil *et al. Journal of Big Data* (2023) 10:108

Page 11 of 17

from 99 percent to 99.85 percent. Furthermore, the minimal accuracy, recall, and F1-score values are 94 percent, 93 percent, and 96 percent, respectively. The maximum accuracy, recall and F1-score values are 100 percent, 99 percent, and 99 percent, respectively. Figure 3 depicts the performance of suggested architecture with other algorithms.

The suggested design is compared to some current architecture using the same dataset. The accuracy of song genre categorization is evaluated using VGG16, Alexnet, and Recurrent Neural Network (RNN), as well as a proposed architecture, as shown in Table 2 and Fig. 4.
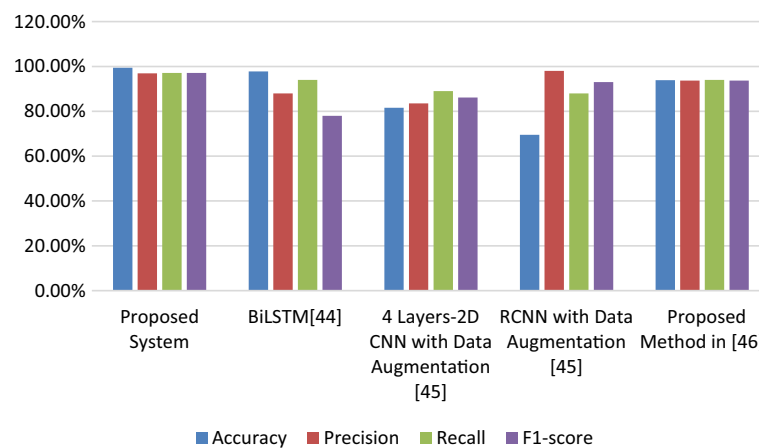
Fig. 3 Performance parameters comparison for GTZAN dataset using proposed architecture

Table 2 Accuracy of music genre categorization in the GTZAN dataset using various topologies

	Proposed architecture	VGG16	Alexnet	RNN
Ours	99.41	90.93	94.55	91.58
Yang et al. [45]	92.5	88.7	88.8	92.5
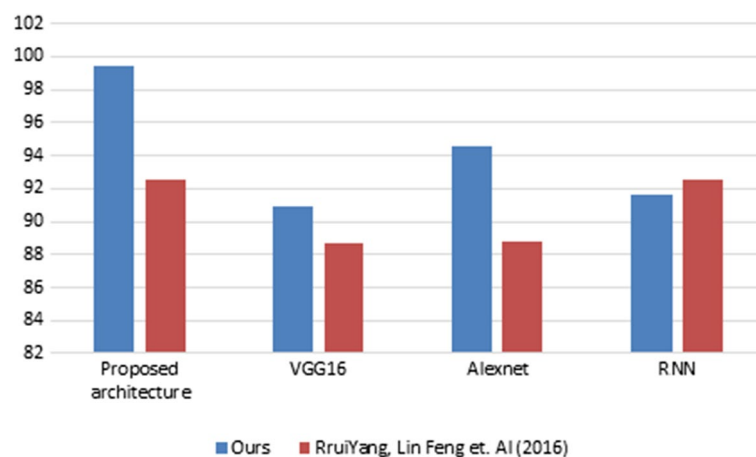
Fig. 4 Comparative analysis of classification accuracy GTZAN dataset using different architectures

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 12 of 17

On the Indian classical taal (rhythm), the planned method is also verified. In Indian classical music, the tabla and dholak are the two main drumming instruments. The tabla is a duo of drums of various woods and sizes that are played together by tapping with the hands to produce a rhythmic sound. The sounds are then arranged in a variety of rhythm patterns to match musical performances. The suggested system architecture is also verified on the Indian classical taal dataset suggested by [40]. The rhythm patterns are evaluated on the basis of Taal: Ektaal, Rupak, Teentaal, and Zaptaal.

In Fig. 5 the performance parameters are shown graphically. All of the characteristics considered for the classification using the suggested proposed method are higher, according to the graph of performance parameters. The accuracy of rhythm classification is at least 92 percent for each class. Rhythmic similarity is tested by using various architecture like Alexnet, 16-layer deep neural network for CNN: VGG 16, and Recurrent Neural Network with suggested architecture which is shown in Table 3. For the classification of taals as compared to other machine learning algorithm mentioned in, proposed architecture gives best performance. Work done by [41], the proposed BP-Model and SP-Model gave less accuracy as compared to proposed method in this paper for HMR dataset.
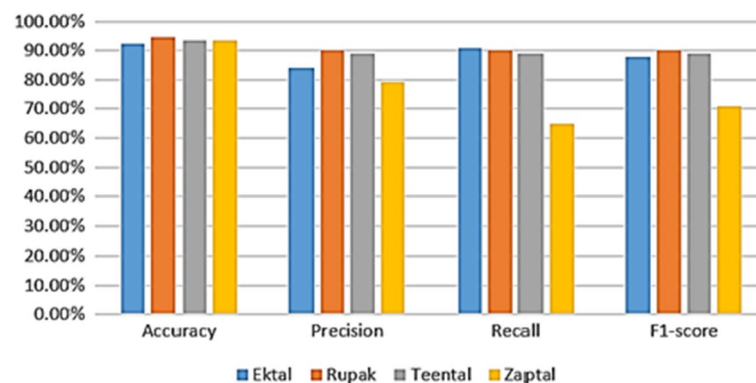
Fig. 5 Performance parameters comparison for the classification of Indian rhythms using proposed architecture

Table 3 Classification accuracy of Indian beats using various designs

	Proposed architecture	SVM	Naïve Bayes	Decision tree	KNN
		Hettiarachchi et al. [46]			
Accuracy	93.44%	48.38%	48.38%	51.61%	38.70%

Table 4 The result compared with other similar method

Method	Accuracy (%)
Proposed method	93.44
Proposed method [41]	63
Data mining method [46]	44.51
Dynamic frame analysis and SVM [49]	98
Entropy and GMM [47]	69.65

Table 5 Comparison of the PSA with existing algorithms based on accuracy on the IMG dataset

IMG dataset genre	PSA (%)	VGG16 (%)	Alexnet (%)	RNN (%)
Bollypop	98.73	97.92	95.23	87.49
Carnatic	99.16	97.16	91.24	89.30
Ghazal	98.41	97.20	92.21	92.61
Semiclassical	99.20	98.64	89.32	84.98
Sufi	99.16	98.28	94.31	82.34

Table 6 Evaluation of Indian music genre per class on the proposed model

	Precision	Recall	F1-Score	Accuracy
IMG dataset				
Bollypop	97%	96%	97%	98.73%
Carnatic	96%	99%	98%	99.16%
Ghazal	96%	96%	96%	98.41%
Semiclassical	99%	97%	98%	99.20%
Sufi	98%	98%	98%	99.16%
HMR dataset				
Ektaal	92.35%	84%	91%	88%
Rupak	94.54%	90%	90%	90%
Teentaal	93.44%	89%	89%	89%
Zaptaal	93.44%	79%	65%	71%
Tabala dataset				
Addhatritaal	0.96	0.97	0.97	99.11
Bhajani	0.96	0.97	0.97	99.11
Deepchandi	0.98	1.0	0.99	99.82
Dadra	0.94	0.99	0.96	99.11
Rupak	0.98	0.95	0.97	99.29
Zaptaal	0.97	0.95	0.96	99.11
Ektaal	0.97	0.94	0.95	98.93
Tritaal	0.98	0.97	0.97	99.11

The suggested architecture gives far better result as compared to other similar work as shown in below Table 4.

Table 5 compares the proposed System Architecture (PSA) classification with those of other algorithms. This study evaluated the proposed system for the IMG dataset along with the existing neural network architectures, VGG16, AlexNet, and RNN. Table 6 shows performance measure on IMG and HMR dataset and Tabala Dataset. HMR dataset has song and music data with 4 taals: Ektaal, Rupak, zaptaal and Teentaal (Tritaal). We have also used Tabala dataset which is exclusively rhythmic dataset with taal on percussion instrument Tabala. As shown in Table 6 we can see that the accuracy and overall performance of Tabala dataset if far better than HMR dataset with the same taal. The deficiency in result is due to same Bol (lyrics) of taal at some places is same for these 4 taalas. The proposed system architecture takes 372 Mb memory for the process of classification. The memory required for the classification is approximately 4 times less than that of the RNN classifier.

Time required to get result (in seconds)
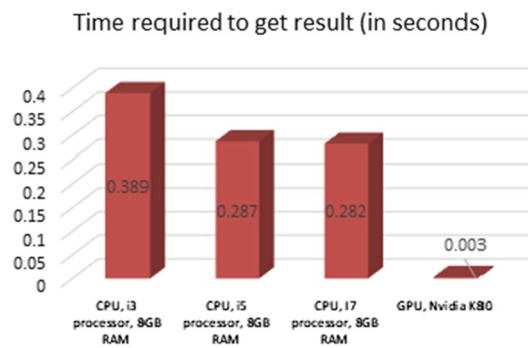
Fig. 6 The time by PSA on different hardware platforms
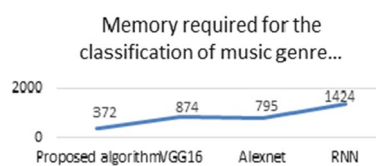
Fig. 7 Memory requirement Graph
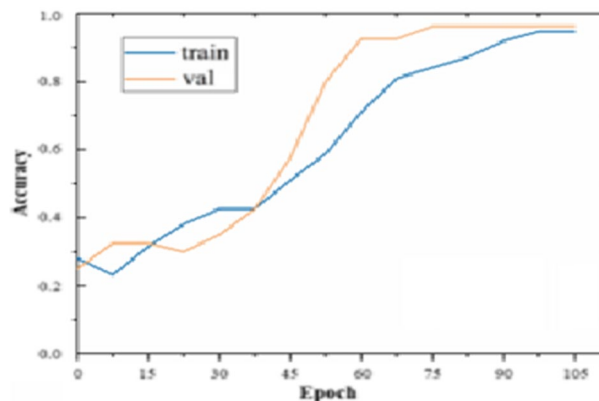
Fig. 8 Learning curve of tabala dataset

The time complexity was tested for numerous CPUs. Typical response times for various hardware systems are shown in Fig. 6. This computational time is much less than time required for music category. Figure 7 shows that the proposed architecture uses 372 MB of memory in the system architecture. Compared with the RNN classifier, this classification requires approximately four times less memory. Figure 8 shows the learning curve of training Vs validation curve.

Conclusion

This research classifies MG using spectrograph feature values from temporal clips of selected music, as well as audio sample not known. The proposed system architecture is evaluated using two datasets: the Indian rhythms and GTZAN. The GTZAN dataset contains genres for western music, which differs significantly from Indian music. We're

also focusing on rhythms in this study. As a result, we employed a rhythmic dataset with a unique pattern to recognize the various and dynamic nature of music. The GTZAN dataset has a classification accuracy of 99.41 percent, while 16-layer CNN, Alexnet, and Recurrent Neural Network have accuracy of 90.93 percent, 94.55 percent, and 91.58 percent, respectively. The suggested system architectural category has an average F1 score of 96.9%, which is significantly higher than current architectures. When evaluated on Indian rhythm, the proposed system design achieves a precision of 93.44 percent, which is higher than earlier architectures. The system's time complexity is quite low. When memory requirements for processing are taken into account, the suggested system architecture outperforms competing categorization techniques. According to the findings, the suggested method gave significant performance as compared to alternative algorithms on the selected GTZAN and Indian rhythms dataset.

Availability of data and materials
The datasets analysed during the current study are available in the [42], https://www.kaggle.com/datasets/carlthome/gtzan-genre-collection and (Ajay Srinivasasmurthy, et al., 2016, dataset https://compmusic.upf.edu/corpus-analysis-hindustani.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Y Panagakis and C Kotropoulos. Music classification by low-rank semantic mappings. Eurasip Journal On Audio, Speech And Music Processing 2013.
2. Li, Tao & Ogihara, Mitsunori & Li, Qi. A Comparative Study on Content-Based Music Genre Classification.2003, 282–289. https://doi.org/10.1145/860435.860487.
3. C Xu, NC Maddage, X Shao, F Cao, Q Tian. Musical genre classification using support vector machines. ICASSP2003, IEEE.
4. R Mellon, D Spaeth, E Theis. Genre classification using graph representations of music, article published, 14 November 2014.
5. P Tyagi, A Mehrotra, S Sharma, and S Kumar. Audio pattern recognition and mood detection system, springer science business media Singapore 2016 M. Pant Et Al. (Eds.), Proceedings of fifth international conference on soft computing for problem solving, advances in intelligent systems and computing.
6. Liu C, Feng L, Liu G, et al. Bottom-up broadcast neural network for music genre classification. Multimedia Tools Appl. 2021;80:7313–31. https://doi.org/10.1007/s11042-020-09643-6.
7. A Ghildiyal, K Singh and S Sharma. Music genre classification using machine learning. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020; pp. 1368–1372, doi: https://doi.org/10.1109/ICECA49313.2020.9297444.
8. Cheng Yu-Huei, Chang Pang-Ching, Nguyen Duc-Man, Che-Nan K. "Automatic music genre classification based on CRNN. Eng Lett. 2021;29(1):36.
9. Foleis JH, Tavares TF. Texture selection for automatic music genre classification. Appl Soft Comput. 2020;89:106127.

10. P Ginsel, I Vatolkin, G Rudolph. Analysis of structural complexity features for music genre recognition. Authorized licensed use limited to: Suny At Stony Brook. Downloaded on October 04, 2020 at 02:07:12 UTC from IEEE Xplore.

11. M Wu, X Liu. A double weighted KNN algorithm and its application in the music genre classification. 2019 6th International Conference on Dependable Systems and Their Applications (DSA).

12. Elbir A, Aydin N. Music genre classification and music recommendation by using deep learning. Electron Lett. 2020;56(12):627–9.

13. Ramírez J, Flores MJ. Machine learning for music genre: multifaceted review and experimentation with audio set. J Intell INF Syst. 2020;55:469–99. https://doi.org/10.1007/s10844-019-00582-9.

14. Pelchat N, Gelowitz CM. Neural network music genre classification. Can J Electr Comput Eng. 2020;43(3):170–3. https://doi.org/10.1109/CJECE.2020.2970144.

15. B Liang and M Gu. Music genre classification using transfer learning. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). 2020; pp. 392-393, doi: https://doi.org/10.1109/MIPR49039.2020.00085.

16. SS Ghosal and I Sarkar. Novel approach to music genre classification using clustering augmented learning method (CALM). Volume 29, Issue 1, 2020.

17. ARS Parmezan, DF Silva, GEAPA Batista. A combination of local approaches for hierarchical music genre classification. Proceedings of the 21st ISMIR conference, Montreal, Canada, October 11–16, 2020.

18. C Yuan, Q Ma, J Chen, W Zhou, X Zhang, X Tang, J Han, S Hu. Exploiting heterogeneous artist and listener preference graph for music genre classification. Poster Session A3: Multimedia search and recommendation & multimedia system and middleware, MM '20, October 12–16, 2020, Seattle, WA, USA.

19. Y Zhuang, Y Chen, J Zheng. Music genre classification with transformer classifier. Proceedings of the 2020 4th international conference on digital signal processing. 2020; doi:https://doi.org/10.1145/3408127.3408137.

20. YH Cheng, PC Chang, CN Kuo. Convolutional neural networks approach for music genre classification "2020 International Symposium on Computer, Consumer and Control (IS3C) | 978-1-7281-9362-5/20/$31.00 ©2020 IEEE | DOI: https://doi.org/10.1109/IS3C50286.2020.00109.

21. N Ndou, R Ajoodha, A Jadhav. Music Genre classification: a review of deep-learning and traditional machine-learning approaches. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-4067-7/21/$31.00 ©2021 IEEE | DOI: https://doi.org/10.1109/IEMTRONICS52119.2021.9422487.

22. Ramírez Castillo, Jaime & Flores, M. Web-Based Music Genre Classification for Timeline Song Visualization and Analysis. IEEE Access. 2021, PP. 1–1. https://doi.org/10.1109/ACCESS.2021.3053864.

23. Goulart AJH, Guido RC, Maciel CD. Exploring different approaches for music genre classification. Egypt Inform J. 2012;13(2):59–63.

24. De Araújo Lima R, de Sousa RCC, Lopes H, Barbosa SDJ. Brazilian lyrics-based music genre classification using a BLSTM network. In: Rutkowski L, Scherer R, Korytkowski M, Pedrycz W, Tadeusiewicz R, Zurada JM, editors. Artificial intelligence and soft computing. ICAISC 2020. Lecture notes in computer science. Cham: Springer; 2020. https://doi.org/10.1007/978-3-030-61401-0_49.

25. RIM Setiadi De, DS Rahardwika, EH Rachmawanto, CA Sari, A Susanto, IUW Mulyono. Effect of Feature Selection on The Accuracy of Music Genre Classification using SVM Classifier. Authorized licensed use limited to: Queens University Belfast. Downloade7d on May 16,2021 at 23:53:53 UTC from IEEE Xplore.

26. A Budhrani, A Patel, S Ribadiya. Music2Vec: Music Genre Classification and Recommendation System "Authorized licensed use limited to: University of Canberra. Downloaded on May 21, 2021 at 06:20:57 UTC from IEEE Xplore.

27. RA Lima de, RCC Sousa de, SDJ Barbosa, HCV Lopes. Brazilian lyrics-based music genre classification using a BLSTM network. arXiv:2003.05377v1 [cs.CL] 6 Mar 2020.

28. GP Oliveira, MO Silva, DB Seufitelli, A Lacerda, MM Moro. Detecting collaboration profiles in success-based music genre networks. Proceedings of the 21st ISMIR Conference, Montr´eal, Canada, October 11–16, 2020.

29. A Nandy, M Agrawal. A novel multimodal music genre classifier using hierarchical attentionand convolutional neural network. arXiv:2011.11970v1 [cs.SD] 24 Nov 2020.

30. R Ozakar, E Gedikli. Music genre classification using novel song structure derived features. Authorized licensed use limited to: Macquarie University. Downloaded on November 13,2020 at 21:41:40 UTC from IEEE Xplore.

31. MS Ahmed, MZ Mahmud, S Akhter. Musical genre classification on the marsyas audio data using convolution NN. 2020 23rd International Conference on Computer and Information Technology (ICCIT). 19–21 December, 2020.

32. C Chen, X Steven. Combined transfer and active learning for high accuracy music genre classification method. 2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE 2021).

33. S Deepak, DBG Prasad. Music classification based on genre using LSTM. Proceedings of the second international conference on inventive research in computing applications (ICIRCA-2020) IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2.

34. Campobello G, Dell'Aquila D, Russo M, Segreto A. Neuro-genetic programming for multigenre classification of music content. Appl Soft Comput J. 2020;94:106488.

35. Domingues, Marcos & Pegoraro Santana, Igor & Pinhelli, Fabio & Donini, Juliano & Catharin, Leonardo & Mangolin, Rafael & Costa, Yandre & Feltrim, Valéria Delisandra. (2020). Music4All: A New Music Database and Its Applications. https://doi.org/10.1109/IWSSIP48289.2020.9145170.

36. Guido RC. Enhancing teager energy operator based on a novel and appealing concept: signal mass. J Frankl Inst. 2018. https://doi.org/10.1016/j.jfranklin.2018.12.00.

37. Guido RC. A tutorial review on entropy-based handcrafted feature extraction for information fusion. Inform Fusion. 2018;41:161–75.

38. Scalvenzi RR, Guido RC, Marranghello N. Wavelet-packets associated with support vector machine are effective for monophone sorting in music signals. Int J Semant Comput. 2019;13(3):415–25.

39. AS Ladkat, AA Date and SS Inamdar. Development and comparison of serial and parallel image processing algorithms. 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-4, doi: https://doi.org/10.1109/INVENTIVE.2016.7824894.

Patil *et al. Journal of Big Data*   (2023) 10:108

Page 17 of 17

40. A Srinivasasmurthy, A Holzapfel, AT Cemgil, X Serra. A generalized Bayesian model for tracking long metrical cycles in acoustic music signals. in Proc. of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Shanghai, China, March 2016.
41. A Srinivasasmurthy, A Holzapfel, X Serra. Informed automatic meter analysis of music recordings. 18th international society for music information retrieval conference, Suzhou, China, 2017.
42. Tzanetakis G, Cook PR. Musical genre classification of audio signals. IEEE Trans Speech Audio Process. 2002;10:293–302.
43. Hongdan W, SalmiJamalib S, Zhengping C, Qiaojuana S, Le R. An intelligent music genre analysis using feature extraction and classification using deep learning techniques. Comput Electr Eng. 2022;100:107978.
44. Vaibhavi M, Krishna PR. Music genre classification using neural networks with data augmentation a make in India creation. J Innov Sci Sustain Technol. 2021;1:21–37.
45. Yang R, Feng L, Wang H, Yao J, Luo S. Parallel recurrent convolutional neural networks based music genre classification method for mobile devices. IEEE Access. 2020. https://doi.org/10.1109/ACCESS.2020.2968170.
46. HBN Hettiarachchi, LS Lekamge, J Charles. A data mining approach, international conference on advances in computing and technology (ICACT–2020), ISSN 2756–9160 / November 2020.
47. AJH Goulart, CD Maciel, RC Guido, KCS Paulo and IN Silva da. Music genre classification based on entropy and fractal lacunarity. 2011 IEEE International Symposium on Multimedia, Dana Point, CA, USA, 2011, pp. 533-536, doi: https://doi.org/10.1109/ISM.2011.94.
48. Elachkar C, Couturier R, Atéchian T, Makhoul A. Combining reduction and dense blocks for music genre classification. In: Mantoro T, Lee M, Ayu MA, Wong KW, Hidayanto AN, editors. Neural information processing. ICONIP 2021. Communications in computer and information science. Cham: Springer; 2021.
49. SH Chen, SH Chen, RC Guido. Music genre classification algorithm based on dynamic frame analysis and support vector machine. 2010 IEEE International Symposium on Multimedia.
50. Kumaraswamy, Balachandra & Poonacha, P.G. Deep Convolutional Neural Network for musical genre classification via new Self Adaptive Sea Lion Optimization. Applied Soft Computing. 2021, 108. 107446. https://doi.org/10.1016/j.asoc.2021.107446.

Publisher's Note