International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 27, No. 2, April, 2022, pp. 327-333
Journal homepage: http://iieta.org/journals/isi

Designing of a Novel Neural Network Model for Classification of Music Genre

Swati A. Patil[1,2*], Thirupathi Rao Komati[1]

[1] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur 522502, A.P., India
[2] G. H. Raisoni Institute of Business Management, Jalgaon 425001, India

Corresponding Author Email: 163030033@kluniversity.in

https://doi.org/10.18280/isi.270217

Received: 28 December 2021
Accepted: 14 April 2022

Keywords:
neural network, feature extraction, music genres

ABSTRACT

Music genre classification is an important task that entails classifying music genres based on aural data. Music genre classification is widely used in the field of music information retrieval. Data preparation, feature extraction, and classification are the three primary processes in the proposed system. New neural network is used to classify music genres. To categorize songs into respective music genres, the proposed system leverages feature values from spectrograms created from slices of songs as input to a proposed system architecture. Extensive tests on the GTZAN dataset demonstrate the efficacy of the proposed approach in comparison to existing methods. The proposed system architecture is also tested on Indian rhythms. This paper consists of the comparison of proposed system architecture with existing algorithms.

1. INTRODUCTION

Great amount of data is generated nowadays on various platforms like YouTube, WhatsApp, Instagram, Twitter etc. Since this data is very high in terms of bytes it is very important to classify the data so that user can search or recommend easily. Our work is basically focusing on music data. Music can be processed or analyzed like audio can be speech, instrumental songs, unplugged songs, songs with music. To process the audio or music signal in computer, it has to be converted into binary format. The audio data is analog signal so analog to digital converter converts it into digital format. The audio signal is represented by the spectrogram or waves.

In this paper CNN as a deep neural network is used. Machine learning algorithms behaves poorly towards images, audios and unstructured data. However, CNN as DNN reduces the images in such a way that it is easy to process.

Users may now listen to music from anywhere and at any time thanks to the advent of online music streaming platforms. They may use these sites to search for millions of songs as per [1]. Music genre categorization has recently piqued the interest of researchers working on recommendation systems, instrument identification, and track separation. Major advancements in music genre classification have recently been made. Music genre classification is an important task that serves as the first stage in developing a recommendation system. Music genre categorization is critical for many real-world applications, such as efficient searching, retrieval, and recommendation [2], due to the large amount of music archives. The main purpose of this paper is to use feature values obtained from music time slices to construct music genre categorization and music recommendation. There are primarily two processes in the categorization of music genres: feature extraction and model construction. The process of identifying individual properties from audio recordings is known as feature extraction [3-5]. Features derived from the music include zero crossing rate, spectral centroid, spectral roll-off, spectral bandwidth, spectral contrast, and Mel-frequency cepstral coefficients [6]. The feature extraction module extracts the most relevant data from the raw music data, which has an impact on the classifier's performance and design. Extracting characteristics of music voice discrimination and speech recognition requires more effort than extracting features from music signals [7]. The Librosa python library [8] is used to acquire all of these functionalities. The features are supplied into the model once the feature values are extracted from spectrograms. CNN is one of the most useful deep learning algorithms, and it was employed in this work. The songs' spectrograms are broken up into pieces and sent into the Librosa library, which extracts the characteristics. To categorize the music, these characteristics are given into the CNN model. The GTZAN dataset was used to compare the performance outcomes of music genre categorization. This paper [9] employs multi-label categorization. The music information retrieval (MIR) community uses deep learning as a more efficient method [10].

The method of collecting information from songs is known as Music Information Retrieval (MIR). In MIR various type of information can be retrieved like genre recommendation, genre classification, fingerprinting, query by humming, musical instrument identification or classification. Deep learning has a number of advantages, one of which is the ability to apply it to large datasets. The various CNN models include VGG16, Alexnet, and Inceptionv3. The key benefit of these models is that they decrease overfitting and improve classification performance. In music genre categorization, neural networks such as the RNN and the Convolutional Neural Network have been used. The research [11] uses an independent RNN to categorize the songs from the GTZAN dataset into genres. Unlike other popular units like the Gated Recurrent Unit, this strategy develops long-term relationships (GRU). Two classifiers termed LSTM and SVM [12]. After collecting features from audio recordings, use a sum rule to combine these classifiers. To categorize the music [13] offers a 1-dimensional CNN. A sliding window is used to separate the music signals in this system. A single audio clip is separated into pieces, which are then combined to form final predictions. Many research on music genre classification have been undertaken in recent years. Deep learning and machine learning approaches are employed in the categorization of music genres. The majority of techniques are based on deep learning. Senac et al. [14] employ spectrograms as input to the CNN model, and the spectrograms are used to extract eight major characteristics as well as three supplementary features such as atonality, dynamics, and timbre. Mirtoolbox [15] was used to capture all of these characteristics. On both the GTZAN and Extended ballroom datasets, Yu et al. [16] utilized three different attention models. The attention mechanism outperforms a system based on encoder-decoders. Both reading and vision-related tasks rely on the attention process. In NMT, it aids in the memorization of lengthier sequences. Because SLA performs poorly, the parallelized attention model has outperformed the serialized attention model. Calculating attention scores and normalizing the attention mechanism to produce attention probabilities are the two phases in the attention mechanism.

Yang et al. [17] uses spectrograms of songs as input to extract features using a concurrent recurrent convolutional neural network. In this system, the CNN and the Bi-RNN are employed in tandem. The music is shown using the STFT. The spectrograms are analysed using CNN for feature extraction and RNN for temporal feature extraction. The benefit of adopting Bi-RNN is that it preserves sequential information that would otherwise be lost during CNN training. A SoftMax function is used to categorize the songs once they have been fully joined. Spectrograms were utilized as input to the CNN model by Despois [18]. The genres Hardcore, Dubstep, Electro, Classical, Rap, and Soundtrack are utilised in this article. First, the songs are transformed to spectrograms, and then the spectrograms are divided into pieces. The tracks are divided into 2.56 seconds each. To categorise the songs into their various musical genres, four CNN layers, a fully connected layer, and a softmax function are employed. Foleis et al. [19] advocated using k-means for texture selection. The main goal of this technology is to recognise sound textures inside recordings. Extracting textures from songs is an excellent way to save storage computational stress.

Liu et al. [20] propose an architecture for broadcast modules that includes inception blocks, transition layers, and decision layers. The broadcast module's main purpose is to retain all extracted characteristics at higher levels of layers so that decision layers may make predictions based on them. Mobile phones and other gadgets have been used to test this strategy. To overcome the challenge of music genre categorization, Scarpiniti et al. [21] adopts a stacked auto-encoder architecture. The approach that extracts 57 characteristics from the music signal. Pelchat et al. [22] utilises CNN to categorise tunes as well. It uses Sound Exchange [23] to convert music data into spectrograms, and it converts music data into stereo channels into mono channels. They utilised a collection of 1880 music tunes as a dataset. Bangare et al. demonstrated excellent study employing machine learning approaches in their publication [24-29]. Sunil et al. [30] and Joseph et al. [31] conducted research using machine learning and deep learning approaches. Awate et al. [32] discussed CNN research for Alzheimer's disease. Mall et al. [33] shown the use of Machine learning for disease detection etc. The work in Pande & Chetty [34] presents a comprehensive assessment of capsule network-based techniques used in image, speech, and signal processing. The authors [35, 36] employed capsule network for effective leaf retrieval.

The suggested system's goal is to categorize music genres. The categorization is based on information taken from the songs' spectrograms. The suggested system method will be extremely beneficial in a variety of fields, including movies, education, and so on. For better user recommendations, several online music streaming systems, such as Spotify and iTunes, employ music genre classification. The following is a breakdown of the paper's structure. Section 2 delves into the proposed system in depth. Section 3 discusses the results analysis, which includes a comparison of the proposed system to existing approaches. The same technique is used to classify Indian cultural songs, and the findings are presented in the same session. Section 4 comes to a close with a conclusion.

2. PROPOSED SYSTEM ARCHITECTURE

Figure 1 depicts a schematic representation of the proposed system architecture. The suggested architecture technique is divided into three sections: a compression, decompression and filtered CNN. All of the convolutions are executed with suitable padding, with the goal of both exploiting characteristics from the input and lowering its resolution by utilizing adequate stride at the conclusion of each step. The proposed architecture design is partially comparable to the widely used VNet paradigm, however there are notable distinctions.

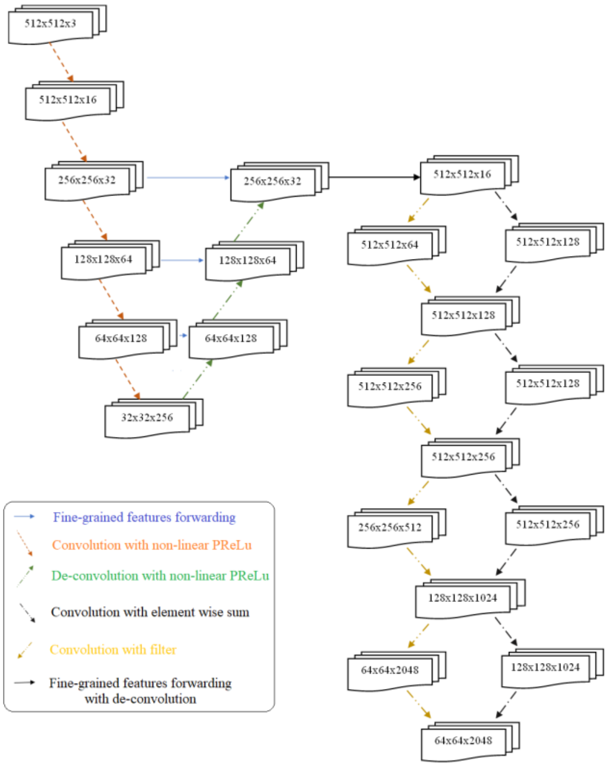

Figure 1. Proposed system architecture

The spectrogram is calculated from the given music. Then by using image processing algorithm different features have been calculated such as zero crossing rate, spectral centroid, spectral roll-off, spectral bandwidth, spectral contrast, and

Mel-frequency cepstral coefficients. Now these features get passed through successive steps along the compression path, and this is accomplished using a convolutional layer with a size of 2×2 and a stride of 2. Because, the second process only evaluates non-overlapping 2 2 patches and extract features, the size of the resultant feature maps is reduced. In our technique, we substituted max-pooling layers with convolutional layers, which achieve the same goal as pooling layers. We doubled the amount of feature maps using these convolutional procedures. This is related to the method's construction as a residual framework, as well as the fact that the number of feature channels doubles at each step of the compression path.

Using convolutional layers instead of pooling layers leads in a lower memory footprint for the network during the training process. The features were down-sampled during the pooling procedure, but we wanted to save as many as feasible. The benefits of employing convolutional layers instead of pooling layers in our suggested technique are that it allows us to analyse inputs at a higher resolution and identify tiny features, as well as collect more contextual information by extending the perspective of input data. The down-sampling stage reduces the size of the input while extending the receptive field of the characteristics being evaluated in the next layers of the network. The number of characteristics analysed by each phase in the left half of the network is two times larger than in the previous layer.

The right component of the network enhances the spatial support of the lower resolution feature maps and extracts features in order to assemble and gather the necessary information for the output of a two-channel segmentation map. The last convolutional layer, with a kernel size of 11, computes the two feature maps and provides output that is comparable in size to the input data. In addition, in this layer, we employed the sigmoid function to turn these two feature maps into probabilistic segmentation maps of the background and foreground regions. In contrast to the Vnet architecture, a deconvolutional operation was performed after each phase of the right part of the network, followed by one to three convolution layers. This comprises half of the 5*5 kernels that were used to expand the size of the input in the previous layer.

In the convolutional stages of the right part of the network, we also used residual functions to train residual functions, as we did in the left section. The extracted characteristics were then passed by horizontal linkages from the early phases of the left side of the network to the right segment, as illustrated in Figure 1. As a result, we were able to increase the quality of final contour prediction by obtaining fine-grained information that would have been overlooked during the compression step otherwise. The model's convergence time has also been enhanced as a result of these linkages.

The ReLU function's strength comes from an army of ReLUs, not from the function itself. This is why a neural network with only a few ReLUs would not produce sufficient results; instead, the network will require a large number of ReLU activations in order to generate a full map of points. Rectified linear units (ReLU) combine in multi-dimensional space to build complex polyhedral along class borders. ReLU does not activate all neurons in neural network at a time, it also helps to improve the training of neural network, no complex computation of gradient, no complex computations like multiplication, division, exponentials.

Input: training data $\{X_{train}, y_{train}\}$, a convex loss function $L(\mathbf{y}, \hat{\mathbf{y}})$, CNN configuration, hyperparameter for low-dimensional binary filter s and m.

Proposed Algorithm:

```
Output: Compressed CNN model

Initialize proxy Features $\{F_1, \ldots, F_m\}$ and $\{\mathbf{Q}^t\}_{t=1}^{Cout}$ for each convolution layer $l$ based on CNN configuration and $s$ and $m$
for iteration =1 to maxIter do
    Get a minibatch of training data $\{X, y\}$
    for $l$=1 to $L$ do
        Obtain low-dimensional binary filters
        $\{\mathbf{B}_1, \ldots, \mathbf{B}_m\}$ according to (7)
        Obtain $\{\mathbf{P}^t\}_{t=1}^{Cout}$ for each convolution filter $t$ ac-
    end for
    Perform standard forward propagation except that convolution operations are defined in Proposition 1
Compute the loss $L(\mathbf{y}, \hat{\mathbf{y}})$
Perform standard De-convolution with non-linear PRelu except that gradients for $\{F_1, \ldots, F_m\}$ and $\{\mathbf{Q}^t\}_{t=1}^{Cout}$ are computed respectively as in (11) and (12)
Perform Fine-grained feature forwarding with de-convolution for proxy variables $\{F_1, \ldots, F_m\}$ and $\{\mathbf{Q}^t\}_{t=1}^{Cout}$ using any popular optimizer
end for

Prediction
Input: test data $\mathbf{X}_{test}$, Trained compressed CNN
Output: predicted labels $\hat{\mathbf{y}}_{test}$;
```

3. RESULTS AND DISCUSSION

The Proposed algorithm is tested on the GTZAN dataset with 390 songs from each class. The confusion matrix clearly indicate that the classic diagonal of the confusion matrix Figure 2 is having dominant values. Confusion matrix also indicates that the misclassification chances get drastically reduced using proposed architecture.

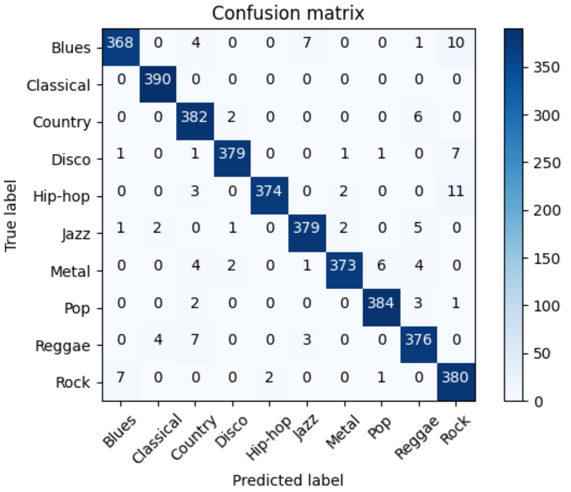

Figure 2. Confusion matrix of proposed algorithm on GTZAN dataset

From the confusion matrix the performance parameters have been calculated. Standard mathematical formulas have been used to calculate the Accuracy, precision, recall and F1-score which is then tabulated in Table 1. The values of False positive rate (FPR) and true positive rate (TPR) are calculated by:

FPR=FP/TN+FP

TPR=TP/TP+FN

Table 1. Performance parameters of proposed algorithm on GTZAN dataset

	Accuracy	Precision	Recall	F1-score
Blues	99.21%	94%	98%	96%
Classical	99.85%	100%	98%	99%
Country	99.26%	98%	95%	96%
Disco	99.59%	97%	99%	98%
Hip-hop	99.54%	96%	99%	98%
Jazz	99.44%	97%	97%	97%
Metal	99.44%	96%	99%	97%
Pop	99.64%	98%	98%	98%
Reggae	99.15%	96%	95%	96%
Rock	99.00%	97%	93%	96%

It clearly shows that the accuracy for classification of Classical songs is the maximum while accuracy for classification of Rock songs is the minimum. As the accuracy of the classification of songs genre is ranging from 99% to 99.85%, we can easily say that the proposed architecture for the classification of songs genre is reliable. Also, minimum value of precision, recall and F1-score is 94%, 93% and 96% respectively. Whereas maximum value of precision, recall and F1-score is 100%, 99% and 99% respectively. Plot in Figure 3 shows the interclass performance parameter values for GTZAN dataset.

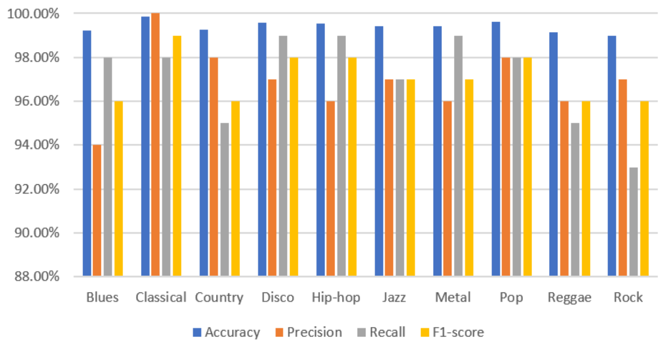

Figure 3. Comparison of the performance parameters for the classification of songs genre of GTZAN dataset using proposed architecture

Table 2. Accuracy of classification of songs genre of GTZAN dataset using different architectures

	Proposed architecture	VGG16	Alexnet	RNN
Blues	99.21%	83.13%	96.41%	96.18%
Classical	99.85%	96.45%	98.24%	92.52%
Country	99.26%	98.35%	95.14%	93.25%
Disco	99.59%	73.45%	96.39%	88.96%
Hip-hop	99.54%	82.42%	86.38%	98.23%
Jazz	99.44%	99.53%	93.15%	79.41%
Metal	99.44%	98.41%	95.18%	87.63%
Pop	99.64%	98.92%	95.32%	91.28%
Reggae	99.15%	81.24%	95.36%	92.53%
Rock	99.00%	97.43%	93.92%	95.78%

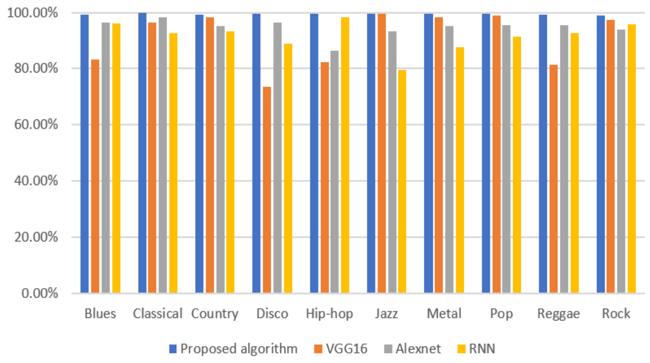

Figure 4. Comparison of classification of songs genre of GTZAN dataset using different architectures

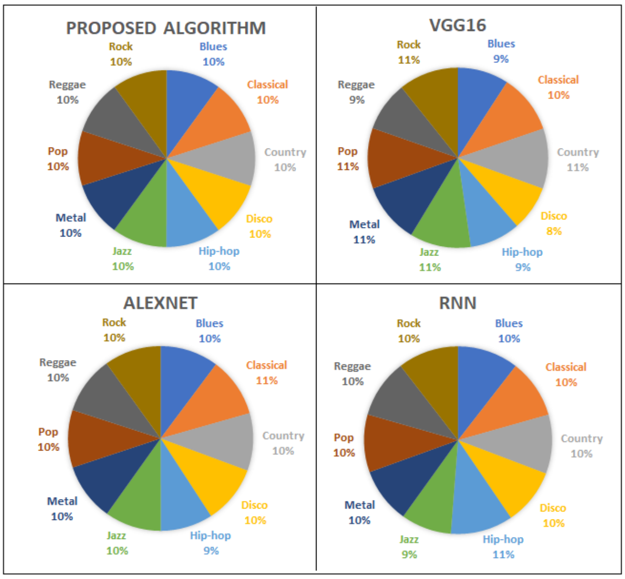

Figure 5. Comparison of Percentage stability of classification of songs genre of GTZAN dataset using different architectures

By using same dataset proposed architecture is tested with some existing architecture. The accuracy of classification of songs genre is tested by using VGG16, Alexnet and RNN along with proposed architecture which is as tabulated in Table 2.

Figure 4 shows the interclass accuracy of the different classes. Here it is noticed that the accuracy of every class for classification is highest than other algorithms. Percentage stability of different algorithms is shown in Figure 5 which indicates the class-wise stability while looking for whole classification system. As here 10 classes are present the stability index should be 10% for each class. Proposed system architecture is having 10% stability index for each class. For classification using VGG16, Alexnet and RNN is unstable. As the accuracy of all the classes is not in the same range the stability index varies from 8% to 11% range. The proposed system architecture is also tested on the Indian classical taal, HMR dataset [37]. In Hindustani Music there are basic 108 taals. Ektaal, Teental, Jhaptal, Rupak are the rhythmic patterns. To understand the rhythm taal are divided into different blocks like Teentaal is of 16 bits having four blocks with four bits, Ektaal is of 12 bits having 6 blocks with two bits, Rupak is if 7 bits which assymtric.in the block size of three, two, two. Since music is very random in nature it is highly challenging to classify music as per genre and get best accuracy [38]. The table and pakhavaj are the two principal percussion instruments used in Indian classical music. The table is a pair of drums of varying sizes and woods that are played simultaneously by tapping them with the hands in various ways to generate diverse sounds. After that, the sounds are put together in varied rhythm patterns to complement musical performances. Ektaal, Rupak, Teentaal and Jhaptaal are the rhythm patterns. The performance parameters are tabulated in Table 3.

Table 3. Performance parameters of proposed algorithm for classification of Indian rhythms

	Accuracy	Precision	Recall	F1-score
Ektaal	92.35%	84%	91%	88%
Rupak	94.54%	90%	90%	90%
Teentaal	93.44%	89%	89%	89%
Jhaptaal	93.44%	79%	65%	71%

The graph as shown in Figure 6 of performance parameters indicates that all the parameters calculated for the classification using proposed architecture are superior. Accuracy of the classification of rhythms is for every class is at least 92%.

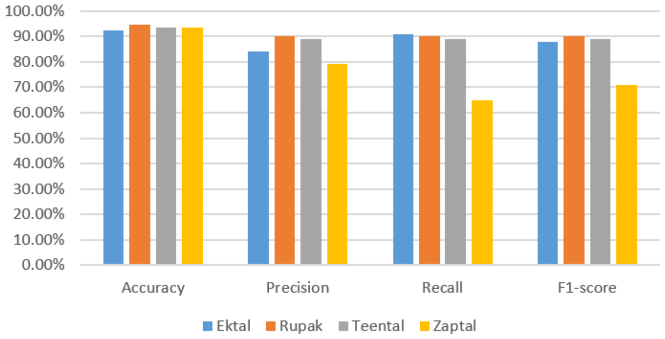

Figure 6. Comparison of the performance parameters for the classification of Indian rhythms using proposed architecture

Indian rhythms are also tested by using VGG16, Alexnet and RNN along with proposed architecture which is as tabulated in Table 4 where Alexnet gives very poor performance while proposed architecture gives best performance for the classification of Indian rhythms. The comparative graph Figure 7 gives better understanding of accuracy of class-wise Percentage stability of different algorithms is shown in Figure 8 which indicates the class-wise stability while looking for whole classification system. As here 4 classes are present the stability index should be 25% for each class. Proposed system architecture is having 25% stability index for each class. For classification using VGG16, Alexnet and RNN is unstable. As the accuracy of all the classes is not in the same range the stability index varies from 23% to 28% in case VGG16, while is varies from 19% to 31% in case of Alexnet and RNN.

Percentage stability of different algorithms is shown in Figure 7 which indicates the class-wise stability while looking for whole classification system. As here 4 classes are present the stability index should be 25% for each class. Proposed system architecture is having 25% stability index for each class. For classification using VGG16, Alexnet and RNN is unstable. As the accuracy of all the classes is not in the same range the stability index varies from 23% to 28% in case VGG16, while is varies from 19% to 31% in case of Alexnet and RNN.

Table 4. Accuracy of classification of Indian rhythms using different architectures

	Proposed architecture	VGG16	Alexnet	RNN
Ektaal	92.35%	83.35%	63.12%	62.59%
Rupak	94.54%	84.73%	97.29%	98.41%
Teentaal	93.44%	88.24%	98.63%	79.42%
Jhaptaal	93.44%	99.42%	59.51%	81.43%

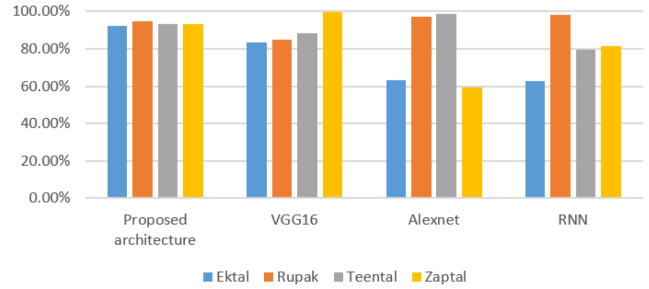

Figure 7. Comparison of classification of Indian rhythms using different architectures

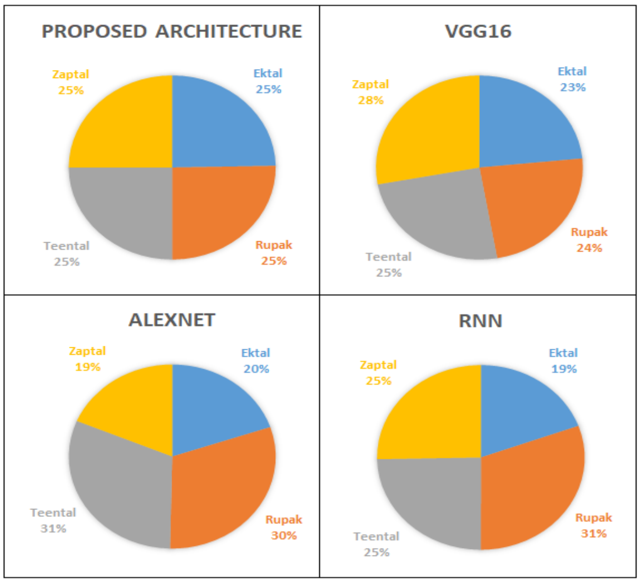

Figure 8. Comparison of percentage stability of classification of Indian rhythms using different architectures

According to Ladkat et al. [39] the suggested system is tested on several processors to determine its time complexity. On different processors, the time complexity is tested. Table 5 shows the average time it takes to receive a result on various hardware platforms.

Table 5. Time complexity of the proposed system architecture on different hardware platforms

Platform	Time required to get result (in seconds)
CPU, i3 processor, 8GB RAM	0.389
CPU, i5 processor, 8GB RAM	0.287
CPU, I7 processor, 8GB RAM	0.282
GPU, Nvidia K80	0.003

When using a CPU, such as an i5, or an i7, the time complexity is almost identical, but when the system is evaluated on a GPU, the time necessary to obtain the findings is significantly different.

4. CONCLUSIONS

Music genre classification is carried out in this study utilizing spectrogram feature values from time slices of songs, as well as an unknown audio clip classified using a majority vote approach. The proposed system architecture is tested on two datasets viz., GTZAN dataset and Indian rhythms. The average accuracy of classification of GTZAN dataset is 99.41% while VGG16, Alexnet and RNN is having accuracy 90.93%, 94.55% and 91.58% respectively. The average F1 score value for the classification using proposed system architecture is 96.9% which is too higher than existing architectures. When the proposed system architecture is tested on Indian rhythm is gives 93.44% accuracy which is better compared to the existing architectures. The experimental results reveal that the suggested system outperforms alternative approaches on the GTZAN dataset as well as Indian rhythms.

REFERENCES

[1] Elbir, A., Çam, H.B., Iyican, M.E., Öztürk, B., Aydin, N. (2018). Music genre classification and recommendation by using machine learning techniques. In 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-5. https://doi.org/10.1109/ASYU.2018.8554016

[2] Corrêaa, D.C., Rodriguesa, F.A. (2016). A survey of symbolic-based music genre classification. Expert Systems with Applications, 60: 190-210. https://doi.org/10.1016/j.eswa.2016.04.008

[3] Mierswa, I., Morik, K. (2005). Automatic feature extraction for classifying audio data. Machine Learning, 58(2): 127-149. https://doi.org/10.1007/s10994-005-5824-7

[4] Tzanetakis, G., Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5): 293-302. https://doi.org/10.1109/TSA.2002.800560

[5] Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, pp. 1096-1103. https://doi.org/10.1145/1390156.1390294

[6] Karatana, A., Yildiz, O. (2017). Music genre classification with machine learning techniques. In 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, pp. 1-4. https://doi.org/10.1109/SIU.2017.7960694

[7] Li, T., Ogihara, M., Li, Q. (2003). A comparative study on content-based music genre classification. In Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 282-289. https://doi.org/10.1145/860435.860487

[8] McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O. (2015). Librosa: Audio and music signal analysis in python. In Proceedings of the 14th Python in Science Conference, pp. 18-25

[9] Prabhu, Y., Kag, A., Gopinath, S., Dahiya, K., Harsola, S., Agrawal, R., Varma, M. (2018). Extreme multi-label learning with label features for warm-start tagging, ranking & recommendation. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 441-449. https://doi.org/10.1145/3159652.3159660

[10] Choi, K., Fazekas, G., Cho, K., Sandler, M. (2017). A tutorial on deep learning for music information retrieval. arXiv preprint arXiv:1709.04396. https://arxiv.53yu.com/abs/1709.04396

[11] Wu, W., Han, F., Song, G., Wang, Z. (2018). Music genre classification using independent recurrent neural network. In 2018 Chinese Automation Congress (CAC), Xi'an, China, pp. 192-195. https://doi.org/10.1109/CAC.2018.8623623

[12] Fulzele, P., Singh, R., Kaushik, N., Pandey, K. (2018). A hybrid model for music genre classification using LSTM and SVM. In 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, pp. 1-3. https://doi.org/10.1109/IC3.2018.8530557

[13] Allamy, S., Koerich, A.L. (2021). 1D CNN Architectures for Music Genre Classification. In 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, pp. 1-7. https://doi.org/10.1109/SSCI50451.2021.9659979

[14] Senac, C., Pellegrini, T., Mouret, F., Pinquier, J. (2017). Music feature maps with convolutional neural networks for music genre classification. In Proceedings of the 15th International Workshop on Content-based Multimedia Indexing, pp. 1-5. https://doi.org/10.1145/3095713.3095733.

[15] Lartillot, O., Toiviainen, P., Eerola, T. (2008). A matlab toolbox for music information retrieval. In Data Analysis, Machine Learning and Applications, pp. 261-268. https://doi.org/10.1007/978-3-540-78246-9_31

[16] Yu, Y., Luo, S., Liu, S., Qiao, H., Liu, Y., Feng, L. (2020). Deep attention based music genre classification. Neurocomputing, 372: 84-91. https://doi.org/10.1016/j.neucom.2019.09.054

[17] Yang, R., Feng, L., Wang, H., Yao, J., Luo, S. (2020). Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices. IEEE Access, 8: 19629-19637. https://doi.org/10.1109/ACCESS.2020.2968170

[18] Despois, J. (2018). Finding the Genre of a Song with Deep Learning—AI Odyssey Part 1.

[19] Foleis, J.H., Tavares, T.F. (2020). Texture selection for automatic music genre classification. Applied Soft Computing, 89: 106127. https://doi.org/10.1016/j.asoc.2020.106127.

[20] Liu, C., Feng, L., Liu, G., Wang, H., Liu, S. (2021). Bottom-up broadcast neural network for music genre classification. Multimedia Tools and Applications, 80(5): 7313-7331. https://doi.org/10.1007/s11042-020-09643-6

[21] Scarpiniti, M., Scardapane, S., Comminiello, D., Uncini, A. (2020). Music genre classification using stacked auto-encoders. In Neural Approaches to Dynamics of Signal Exchanges, pp. 11-19. https://doi.org/10.1007/978-981-13-8950-4_2

[22] Pelchat, N., Gelowitz, C.M. (2020). Neural network music genre classification. Canadian Journal of

Electrical and Computer Engineering, 43(3): 170-173. https://doi.org/10.1109/CJECE.2020.2970144

[23] Norskog, L. (2015). SoX—Sound Exchange. Available: https://linux.die.net/man/1/rec.

[24] Bangare, S.L., Patil, M., Bangare, P.S., Patil, S.T. (2015). Implementing tumor detection and area calculation in MRI image of human brain using image processing techniques. Int. Journal of Engineering Research and Applications, 5(4): 60-65.

[25] Bangare, S.L., Dubal, A., Bangare, P.S., Patil, S.T. (2015). Reviewing Otsu's method for image thresholding. International Journal of Applied Engineering Research, 10(9): 21777-21783. http://dx.doi.org/10.37622/IJAER/10.9.2015.21777-21783

[26] Bangare, S.L., Pradeepini, G., Patil, S.T. (2018). Regenerative pixel mode and tumour locus algorithm development for brain tumour analysis: A new computational technique for precise medical imaging. International Journal of Biomedical Engineering and Technology, 27(1-2): 76-85. https://doi.org/10.1504/IJBET.2018.093087

[27] Bangare, S.L., Pradeepini, G., Patil, S.T. (2017). Neuroendoscopy adapter module development for better brain tumor image visualization. International Journal of Electrical and Computer Engineering, 7(6): 3643. http://dx.doi.org/10.11591/ijece.v7i6.pp3643-3654

[28] Bangare, S.L. (2022). Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images. Neuroscience Informatics, 2(3): 100019. https://doi.org/10.1016/j.neuri.2021.100019

[29] Bangare, S.L., Pradeepini, G., Patil, S.T. (2017). Brain tumor classification using mixed method approach. In 2017 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, pp. 1-4. https://doi.org/10.1109/ICICES.2017.8070748.

[30] Bangare, S., Bangare, M.L., Bangare, P.M., Apare, R.S (2021). Fog computing based security of IoT application. Design Engineering (Toronto), 2021(7): 7542-7549.

[31] Joseph, L.L., Shrivastava, P., Kaushik, A., Bangare, S.L., Naveen, A., Raj, K.B., Gulati, K. (2021). Methods to identify facial detection in deep learning through the use of real-time training datasets management. EFFLATOUNIA-Multidisciplinary Journal, 5(2): 1298 - 1311.

[32] Awate, G., Bangare, S., Pradeepini, G., Patil, S. (2018). Detection of Alzheimers disease from MRI using convolutional neural network with tensorflow. arXiv preprint arXiv:1806.10170. https://arxiv.org/abs/1806.10170.

[33] Mall, S., Srivastava, A., Mazumdar, B.D., Mishra, M., Bangare, S.L., Deepak, A. (2022). Implementation of machine learning techniques for disease diagnosis. Materials Today: Proceedings, 51: 2198-2201. https://doi.org/10.1016/j.matpr.2021.11.274

[34] Pande, S.D., Chetty, M.S.R. (2018). Analysis of capsule network (Capsnet) architectures and applications. J Adv Res Dynam Control Syst, 10(10): 2765-2771.

[35] Pande, S.D., Chetty, M.S.R. (2021). Fast medicinal leaf retrieval using CapsNet. In International Conference on Intelligent and Smart Computing in Data Analytics, pp. 149-155. https://doi.org/10.1007/978-981-33-6176-8_16

[36] Pande, S., Chetty, M. (2019). Bezier curve based medicinal leaf classification using capsule network. International Journal of Advanced Trends in Computer Science and Engineering, 8(6): 2735-42. https://doi.org/10.30534/ijatcse/2019/09862019

[37] Srinivasamurthy, A., Holzapfel, A., Cemgil, A.T., Serra, X. (2016). A generalized bayesian model for tracking long metrical cycles in acoustic music signals. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, pp. 76-80. https://doi.org/10.1109/ICASSP.2016.7471640

[38] Patil, S.A., Rao., K.T. (2021). Machine learning based classification comparison of music genres. Drugs and Cell Therapies in Hematology, 10(1): 1614-1624.

[39] Ladkat, A.S., Date, A.A., Inamdar, S.S. (2016). Development and comparison of serial and parallel image processing algorithms. In 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, pp. 1-4. https://doi.org/10.1109/INVENTIVE.2016.7824894