Educate Elevate Enlighten

JSS Mahavidyapeetha

JSS Science And Technology University

(Established Under JSS Science and Technology University Act No. 43 of 2013)

(Formerly Known as SJCE)

JSS SCIENCE AND TECHNOLOGY UNIVERSITY MYSURU

Subject name: DIGITAL SIGNAL PROCESSING (EC550)

EVENT II AND IV

Bollywood Music Era recognition using sequential neural network

Submitted by

Sl. No.	USN	NAME
1.	01JST18EC105	VARADRAJ YALAMALLI
2.	01JST18EC108	XAVIER FERNANDIS

Submitted to

SHASHIDAR R

Assistant Professor

Department of E & C

JSSSTU, SJCE, Mysore

DEPARTMENT OF ELECTRONICS AND COMMUNICATION

JSS SCIENCE AND TECHNOLOGY UNIVERSITY MYSORE-570006

2020-2021

Educate Elevate Enlighten

JSS Mahavidyapeetha

JSS Science And Technology University

(Established Under JSS Science and Technology University Act No. 43 of 2013)

(Formerly Known as SJCE)

JSS SCIENCE AND TECHNOLOGY UNIVERSITY MYSURU

Subject name: DIGITAL SIGNAL PROCESSING

EVENT II AND IV

Name	**VARADRAJ YALAMALLI**
Roll Number	**53**
Section	**B**
USN	**01JST18EC105**
Submission Date	

Evaluation Component	**Max. Marks**	**Marks Scored**
Presentation skills	**8**	
Documentation	**8**	
Interaction/Viva	**4**	
Total	**20**	

Name of the faculty	**SHASHIDAR R**
Signature of the faculty	
Signature of the student	

Comments:

Educate Elevate Enlighten

JSS Mahavidyapeetha

JSS Science And Technology University

(Established Under JSS Science and Technology University Act No. 43 of 2013)

(Formerly Known as SJCE)

JSS SCIENCE AND TECHNOLOGY UNIVERSITY MYSURU

Subject name: DIGITAL SIGNAL PROCESSING

EVENT II AND IV

Name	**XAVIER FERNANDIS**
Roll Number	**55**
Section	**B**
USN	**01JST18EC108**
Submission Date	

Evaluation Component	**Max. Marks**	**Marks Scored**
Presentation skills	**8**	
Documentation	**8**	
Interaction/Viva	**4**	
Total	**20**	

Name of the faculty	**SHASHIDAR R**
Signature of the faculty	
Signature of the student	

Comments:

TABLE OF CONTENTS

INDEX	CONTENT	PAGE NO.
CHAPTER I	ABSTRACT	5
CHAPTER II	INTRODUCTION	6
CHAPTER III	LITERATURE REVIEW	7
CHAPTER IV	OBJECTIVES	8
CHAPTER V	METHODOLOGY	9
CHAPTER VI	BLOCK DIAGRAM	12
CHAPTER VII	RESULTS	14
CHAPTER VIII	CONCLUSION AND FUTURE WORK	16
CHAPTER IX	REFERENCES	17

Bollywood Music Era recognition using sequential neural network

CHAPTER I:

ABSTRACT

We use the frame level audio features which summarize frequency content within short intervals of time. Comparatively, more recent music information retrieval tasks take advantage of temporal structure in audio spectrograms using deep convolutional and recurrent models. a Convolutional Sequential Neural Network (CNN), is applied to the dataset under a comprehensive set of conditions. These include audio clip length, which is a novel contribution in this work, and previously identified considerations such as dataset split and feature level. Additionally, to showcase the effectiveness of CNN's feature extraction capabilities, we visualize audio samples at the model's bottleneck layer demonstrating that learned representations segment into clusters belonging to their respective artists. Index Terms—artist classification, music, information retrieval, deep learning, convolutional recurrent neural network

CHAPTER II:

INTRODUCTION

The ability to classify music has many applications, from automatic organization of musical databases to music recommendation systems. However, even to a trained human ear, differentiating between similar subgenres of music can be extremely difficult. As a result, there are rarely clear distinctions between similar genres. In particular, subgenres within music are often blurred.

There has been many developments in Music genre recognition ,So in this project we classify Music based on era which may help on the problem of classifying music by subgroups within a specific genre in hopes of gaining insight into the limits of music classification.

CNN is a Deep Learning algorithm which can take an input image as input, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. A CNN has various layers such as Convolutional layers, ReLU layers, Pooling layers and a fully connected layer as shown in figure 6. CNN is widely used for image classification because it does automatic feature extraction using convolution.

CHAPTER III:

LITERATURE REVIEW

Much research has been done on music information retrieval,Vishnupriya S and K Meenakshi [1] have proposed a Neural Network Model to perform the classification. Tzanetakis and Cook [2] pioneered their work on music genre classification using machine learning algorithm. They created the GTZAN dataset which is till date considered as a standard for genre classification. Changsheng Xu et al [3] have shown how to use Convolution Neural Network (CNN) for this task. Matthew Creme, Charles Burlin, Raphael Lenain from Stanford University [4] have used 4 different methods to perform the classification. They have used Support Vector Machines, Neural Networks, Decision Trees and K-Nearest Neighbours methods to perform classification. Tao [5] shows the use of restricted Boltzmann machines and arrives to better results than a generic multilayer neural network by generating more data out of the initial dataset, GTZAN.

After carrying out the above mentioned literature survey, Convolutional Neural Network is used to perform classification and the details of the same are explained in the following sections.

CHAPTER IV:

OBJECTIVES:

The objective of our project is to find a neural network model to take the input of a bollywood song and detect its era. since we are using the audio signal as it is a sequential data we have to use a sequential neural network technique to implement this we have used the RNN model to this project , as it is computationally efficient than the LSTM and and have a higher accuracy than the basic neural network model . While training the RNN we have to find a proper epoch such that the model is properly trained and also computationally efficient . In this project we have to choose a proper activation function. There are many efficient activation functions like Relu , sigmoid , tanh etc .we introduce drop out and other methods to avoid the above problem of overfitting and underfitting and the accuracy and the loss function is calculated on the test set , based on the result we change the model the drop outs , activation function and the hidden layers. After getting a good accuracy we test the result with real time input and test the model .

CHAPTER V:

METHODOLOGY

CREATING AND PREPROCESSING THE DATA SET

The data set has been created using matlab, The mp3 zip file of 5 different era's is been downloaded from various websites and converted into a wav file using the matlab code and trimmed to a length of 1 min . Each wav file is converted into a spectrogram using librosa with the nonequispaced fast Fourier transform length of 2048 ,sampling frequency of 2Khz and plotted with the DB scale and we save the spectrogram in 5 different folders for each era. spectrograms are created for the entire length of each song to form an initial dataset. For the song split, this dataset undergoes a 90/10 stratified-by-era split to create train and test sets respectively. The train set is then split using the same stratified 90/10 split to create train and validation (used for early stopping) subsets. The stratification ensures that each set contains an equivalent number of songs from each era. For the album split,

A short-time Fourier transform is applied to the raw audio signal for every song to create spectrograms. Once created, the frequency scale (f hertz) is transformed into the Mel scale (m mels) using and then scaled (d decibels).These operations are considered standard practices for audio processing and have been shown in prior work to improve performance in classification tasks.

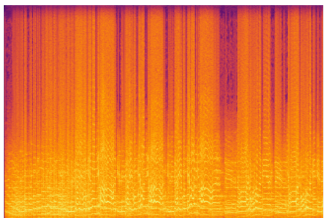

fig 1.1

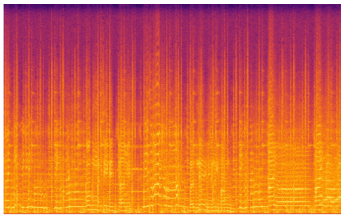

fig 1.2

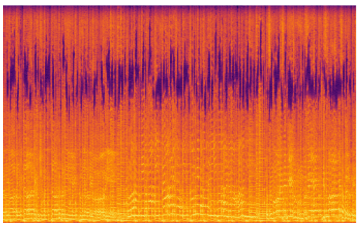

fig 1.3

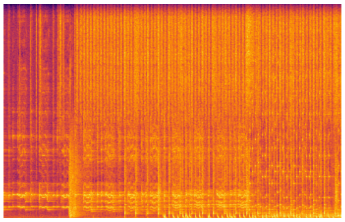

fig 1.4

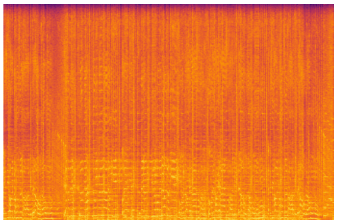

fig 1.5

The above figures show the spectrograms of 5 different era's from the 1960's to 2000's respectively. We can see that there are changes in the music pattern , the instruments, the bass genre and other features as the time progresses . It is wise to classify the music into decades as we can see distinct variations in the music over a decade rather than a year or two. Since the above five figures are different we can classify the era based on this differences in the image of the spectrogram

TRAINING THE DATASET

The source folder is the input parameter containing the images for different classes. We Read the image file from the folder and convert it to the right color format. Resize the image based on the input dimension required for the model.Converted the image to a Numpy array with float32 as the datatype.We Normalize the image array to have values scaled down between 0 and 1 from 0 to 255 for a similar data distribution, which helps with faster convergence.

We use two layers containing filters 32, 64 ,kernel size is 3 which defines An integer or tuple/list of 2 integers, specifying the height and width of the 2D convolution window. Can be a single integer to specify the same value for all spatial dimensions.strides (2,2) An tuple/list of 2 integers, specifying the strides of the convolution along the height and width. Can be a single integer to specify the same value for all spatial dimensions.two dense layers are created with number of classes for 1st layer is 100 clases and activation layer is relu,2nd layer has 5 clases with softmax as activation which gives the probability for given output class.

According to above model is fitted to create a dataset for which the parameters are identified for which we have got 205,082,397 which we have evaluated for the test set with accuracy of 0.95.

CHAPTER VI:

BLOCK DIAGRAM

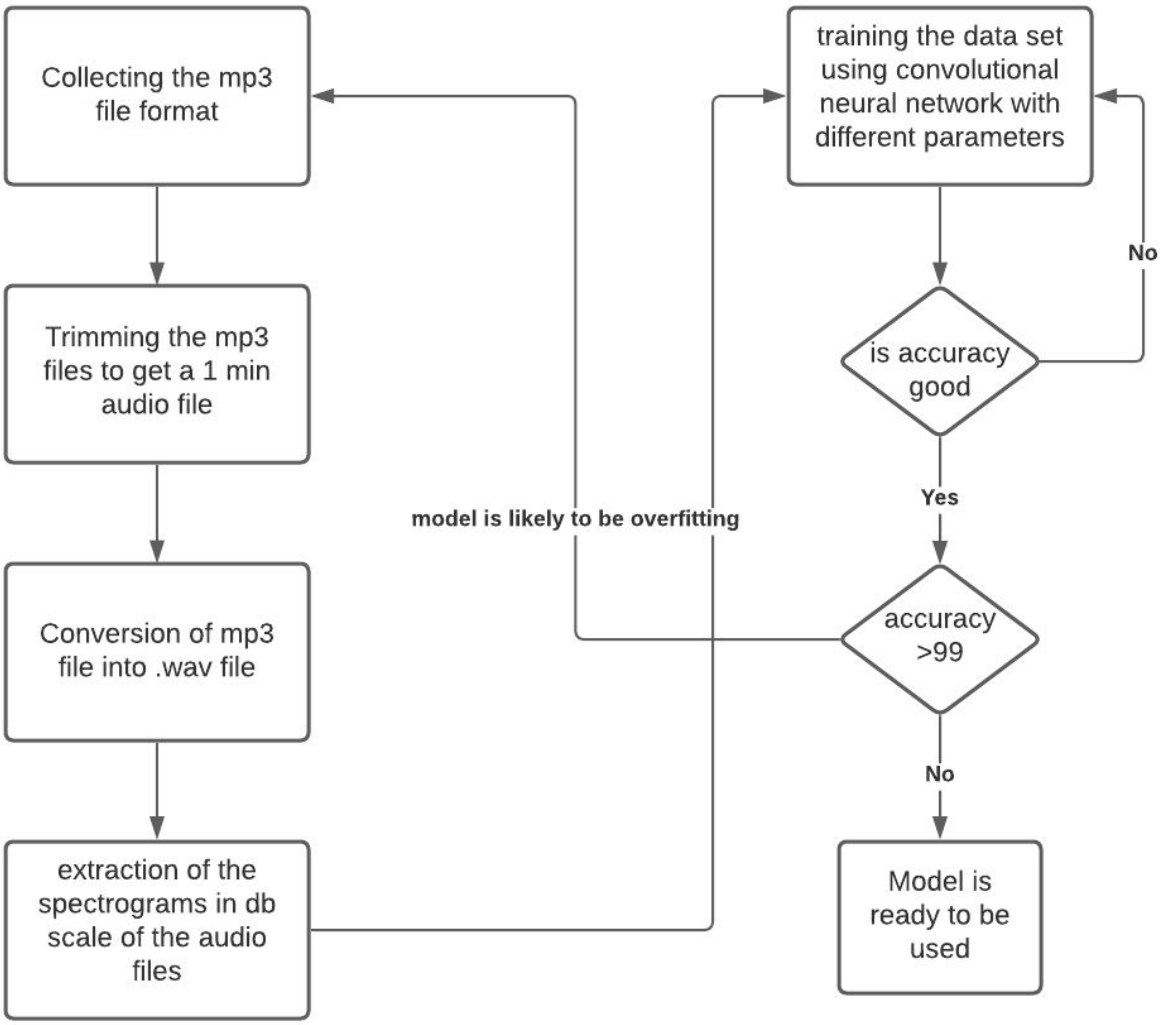

fig 6.1

figure 6.1 is the block diagram on how we did the literature survey and implemented the project . Phase 1 is the preprocessing step where we collect and process the data. The data of bollywood songs has been collected through youtube and it's been processed such that its length is 1 min. later we convert it into spectrograms .We have used the librosa package of python which is an audio processing package Each wav file is converted into spectrogram using librosa with the nonequispaced fast Fourier transform length of 2048 ,sampling frequency of 2Khz and plotted with the DB scale

The second phase is to model a neural network algorithm we have used a recurrent neural network algorithm for our project we trained the data using sklearn library in python , due to the less data available we faced the problem of overfitting and we had to collect more data and the problem of overfitting was solved . We saw that the relu activation function was better to use in our model and we trained out data set into 25 epochs with 2 hidden layer of shape (359,359,32) and (179,179,64) , And we got a satisfactory result from this model on the test set with the accuracy of accuracy of 95.92% .

CHAPTER VII:

RESULTS

The model is being after training the dataset with 25 epochs and the model is built with the 2 hidden layers which uses the relu activation function , with conv2d shape of (359,359,32) with 896 parameters and the second layer conv2d_1 with the output shape of (179,179,64) with 18,496 parameters .

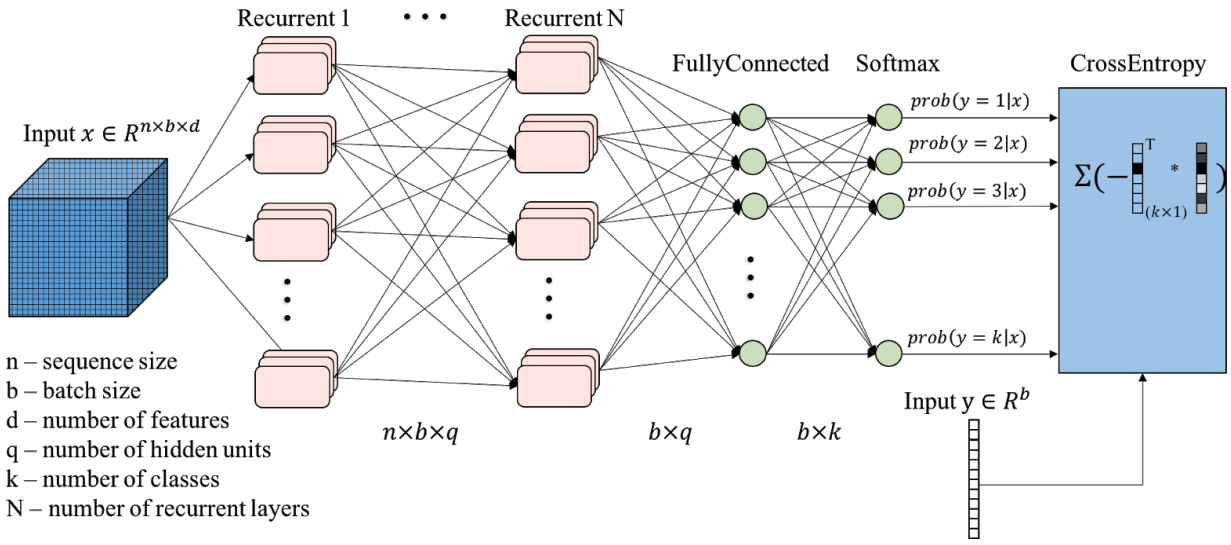

fig 7.1

$$h_t = Wf(h_{t-1}) + W^{(hx)}x_{[t]}$$
$$\hat{y}_t = W^{(S)}f(h_t)$$

- **Total error is the sum of each error at time steps t**

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \frac{\partial E_t}{\partial W}$$

- **Hardcore chain rule application:**

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

fig 7.2

figure 7.1 and fig 7.2 is the mathematical model of convolutional RNN. tells us about the forward and backward propagation and the given input feedback error. We got 20,50,62,500 number of valid training parameters. The model was found to have an accuracy of 95.92% and with the loss of 0.1709.

The model is created for the above specified architecture and number of parameters obtained after each layer. To Check the efficiency and the real time output of the model we have collected spectrograms of 5 test data of different eras. The above is the spectrogram of a 1980's song. We can see that the output that we get is 1980 which is correct. and for all 5 test data the output is verified .hence the model satisfies the required problem statement

CHAPTER VIII:

CONCLUSION AND FUTURE WORK

This work provides the details of an application which performs Music Era Classification using Machine Learning techniques. The application uses a Convolutional Neural Network model to perform the classification. A Mel Spectrum of each track from the local dataset is obtained. This is done by using the librosa package of python. A piece of software is implemented which performs classification of huge databases of songs into their respective era. The extension of this work would be to consider bigger data sets and also tracks in different formats(mp3, au etc). Also, with time the style represented by each era will continue to change. So the objective for the future will be to stay updated with the change in styles of era and extend our software to work on these updated styles. This work can also be extended to work as a music recommendation system depending on the age of the person. Also due to the limited number of data sets on the bollywood music there is a problem of overfitting hence in the future with the increase in the data set the accuracy of the model can be increased.

CHAPTER VIII:

REFERENCES

[1] Musical Genre Classification of Audio Signals George Tzanetakis, Student Member, IEEE, and Perry Cook, Member, IEEE

[2] F. Pachet and D. Cazaly, "A classification of musical genre," in Proc. RIAO Content-Based Multimedia Information Access Conf., Paris, France, Mar. 2000.

[3] G. Tzanetakis and P. Cook, "Multi Feature audio segmentation for browsing and annotation," in Proc. Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, 1999.

[4] K. Choi, G. Fazekas, and M. Sandler, "Explaining deep convolutional neural networks on music classification," arXiv preprint arXiv:1607.02444, 2016.

[5] S. Li, "Content-based classification and retrieval of audio using the nearest feature line method," IEEE Trans. Speech Audio Processing, vol. 8, pp. 619–625, Sept. 2000.

[6] R. N. Shepard, "Circularity in judgments of relative pitch," J. Acoust. Soc. Amer., vol. 35, pp. 2346–2353, 1964.

[7] J. Pierce, "Consonance and scales," in Music Cognition and Computerized Sound, P. Cook, Ed. Cambridge, MA: MIT Press, 1999, pp. 167–185.

[8] J. Nibbeling, "Artist recognition with convolutional recurrent neural networks," 2018.

[9] W. J. Pielemeier, G. H. Wakefield, and M. H. Simoni, "Time-frequency analysis of musical signals," Proceedings of the IEEE, vol. 84, no. 9, pp. 1216–1230, 1996.