**CHAPTER 1**

**INTRODUCTION**

**1.1 Music Generic Classification -**

In today’s world, an individual’s music collection generally contains hundreds of songs, while the professional collection normally contains tens of thousands of music files. Music databases are incessantly gaining reputation in relations to specialized archives and private sound collections. With improvements in internet services and increase in

network bandwidth there is also an increase in number of people accessing the music database.

Dealing with extremely large music databases is exhausting and time consuming. Most of the music files are stored according to the song title or the artist name. This may cause trouble in searching for a song related to a specific genre. The popular music genres are Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae

and Rock.

Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well. This makes music genre classification difficult. Also the definition of music genre has changed over time. For instance, pop songs that were made fifty years ago are different from the pop songs we have today. Fortunately, the progress in music data and its storage has improved considerably over the past few years. Since manually classifying each track of a large music database according to their genre is a tedious task, Machine Learning Techniques to perform Automatic Music Genre Classification are used.

**1.2 Music -**

Music is the art of arranging sounds in time through the elements of melody, harmony, rhythm, and timbre.It is one of the universal cultural aspects of all human societies. General definitions of music include common elements such as pitch (which governs melody and harmony), rhythm (and its associated concepts tempo, meter, and articulation), dynamics (loudness and softness), and the sonic qualities of timbre and texture (which are sometimes termed the "color" of a musical sound). Different styles or types of music may emphasize, de-emphasize or omit some of these elements. Music is performed with a vast range of instruments and vocal techniques ranging from singing to rapping; there are solely instrumental pieces, solely vocal pieces (such as songs without instrumental accompaniment) and pieces that combine singing and instruments. The word derives from Greek μουσική, mousiké, '(art) of the Muses.

Generally, music is known through the cognition of specific arrangements of sounds and the emotional reaction that it engenders. Whether the music is a complex and intellectually demanding orchestral composition or a plaintive folk song, the range of responses can encompass the full range of human emotions. This makes the musical arts a universal expression of human experience.

Music is a vital aspect of culture. It has the ability to invoke age-old traditions or to incite people to social change. Music's ability to connect individuals to social and cultural experience is one of its most important characteristics. Responses to music can often be visceral, and in the twentieth century the global appeal of jazz, blues, folk music, and rock and roll advanced egalitarian values and contributed to loosening sexual mores. In the twenty-first century the dissemination of emotionally charged music through new electronic media provides musicians and composers unprecedented worldwide exposure and influence.



Figure 1.1

Music can be written in several ways. When it is written on a staff (like in the example shown), the pitches (tones) and their duration are represented by symbols called notes. Notes are put on the lines and in the spaces between the lines. Each position says which tone must be played. The higher the note is on the staff, the higher the pitch of the tone. The lower the notes are, the lower the pitch. The duration of the notes (how long they are played for) is shown by making the note "heads" black or white, and by giving them stems and flags.

Music can also be written with letters, naming them as in the solfa "Do, Re, Mi, Fa, So, La, Ti, Do" or representing them by letters. The next table shows how each note of the solfa is represented in the Standard Notation.

**1.3 Different Music Genre -**

A music genre is a conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions. It is to be distinguished from musical form and musical style, although in practice these terms are sometimes used interchangeably.

Music can be divided into genres in varying ways, such as popular music and art music, or religious music and secular music. The artistic nature of music means that these classifications are often subjective and controversial, and some genres may overlap.



Figure 1.2

**1.4 Data Set**

For the completion of this experiment, the entire dataset is downloaded and prepared from kaggle. The selected dataset consists of 9 audio files of different genres such as Blues, Classical, Country, Disco, Hip-hop, Metal, Pop, Reggae, Rock. Here the dataset, consist of audio files thus these audio files are converted in image format using Spectrogram and Histogram and further these generated images are used to train our neural network. We also divided each audio into ten parts so as to increase our dataset. The brief detail of the dataset is that it consists of total 9000 different values for each genres which is further divided into training and testing part in 80:20. Where 80 percent of images are used to train neural network where as 20 percent are used for testing.

|  |  |
| --- | --- |
| **Genre** | **Number of Tracks** |
| **BLUES** | 100 |
| **CLASSICAL** | 100 |
| **COUNTRY** | 100 |
| **DISCO** | 100 |
| **HIPHOP** | 100 |
| **METAL** | 100 |
| **POP** | 100 |
| **REGGAE** | 100 |
| **ROCK** | 100 |

**1.5 Signal**

A signal is an electrical or electromagnetic current that is used for carrying data from one device or network to another.Hence, a signal can be a source of energy which transmits some information. This can easily be represented on a graph.

A signal can be of any type that conveys some information. This signal produced from an electronic equipment, is called as Electronic Signal or Electrical Signal. These are generally time variants.

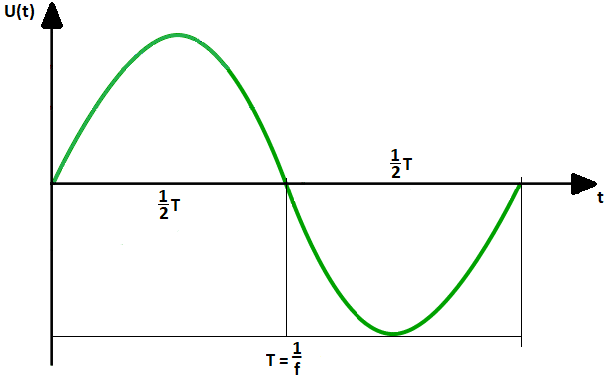


Figure 1.3

**1.6 Time Domain VS Frequency Domain**

|  |  |
| --- | --- |
| **Time domain** | **Frequency Domain** |
| A time domain graph shows how a signal changes over time. | The frequency domain graph shows how much of the signal lies within each given frequency band over a range of frequencies. |
| Time domain is the domain for analysis of mathematical functions or signals with respect to time. | Frequency domain is the domain for analysis of mathematical functions or signals with respect to frequency. |
| The time domain systems tend to use photon counting detectors which are slow but highly sensitive. | The frequency domain systems are relatively inexpensive, easy to develop and use and can provide very fast temporal sampling. |

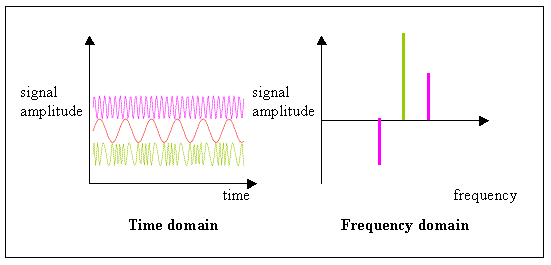


Figure 1.4

**1.7 Fast Fourier Transformation FFT**

Fourier Transform is a mathematical concept that can decompose a signal into its constituent frequencies.FFT algorithm can convert time-domain discrete signal into a frequency-domain.

Fourier transform does not just give the frequencies present in the signal, It also gives the magnitude of each frequency present in the signal. FFT reduces the number of computations needed for a problem of size N from O(N^2) to O(NlogN)

FFT-based features generation is based on the decomposition of the complex signals to smaller transforms. The decomposed signals are combined to find the resulting transform signal. Low frequencies are removed from the obtained signal using FFT and noises are eliminated after the application of the inverse FFT.

The basic idea of the FFT is to apply divide and conquer. We divide the coefficient vector of the polynomial into two vectors, recursively compute the DFT for each of them, and combine the results to compute the DFT of the complete polynomial.

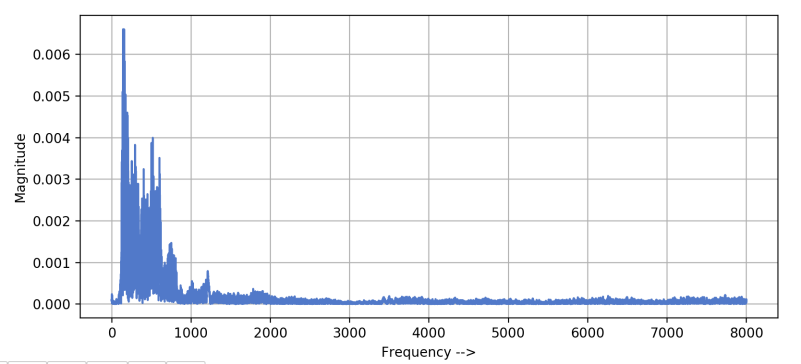


Figure 1.5

**Need of Time-Frequency System**

When we applied FFT to our signal, it gave us only frequency values and we lost the track of time information. Now our system won’t be able to tell what was spoken first if we use these frequencies as features.

To get better features , we need frequency with respect to time.

Visual representation of frequencies of a given signal with time is called Spectrogram.We can get a spectrogram using STFT (Short time fourier transformation)

Examples of time frequency system representation : **Spectrogram , Mel - Spectrogram**

**1.8 Spectrogram**

A spectrogram is a visual way of representing the signal strength, or “loudness”, of a signal over time at various frequencies present in a particular waveform. Not only can one see whether there is more or less energy at, for example, 2 Hz vs 10 Hz, but one can also see how energy levels vary over time. In other sciences spectrograms are commonly used to display frequencies of sound waves produced by humans, machinery, animals, whales, jets, etc., as recorded by microphones. In the seismic world, spectrograms are increasingly being used to look at frequency content of continuous signals recorded by individual or groups of seismometers to help distinguish and characterize different types of earthquakes or other vibrations in the earth.

Spectrograms are basically two-dimensional graphs, with a third dimension represented by colors. Time runs from left (oldest) to right (youngest) along the horizontal axis. Each of our volcano and earthquake sub-groups of spectrograms shows 10 minutes of data with the tic marks along the horizontal axis corresponding to 1-minute intervals. The vertical axis represents frequency, which can also be thought of as pitch or tone, with the lowest frequencies at the bottom and the highest frequencies at the top. The amplitude (or energy or “loudness”) of a particular frequency at a particular time is represented by the third dimension, color, with dark blues corresponding to low amplitudes and brighter colors up through red corresponding to progressively stronger (or louder) amplitudes.

To generate a spectrogram, a time-domain signal is divided into shorter segments of equal length. Then, the fast Fourier transform (FFT) is applied to each segment. The spectrogram is a plot of the spectrum on each segment. The Frame Count parameter determines the number of FFTs used to create the spectrogram and, as a result, the amount of the overall time signal that is split into independent FFTs.

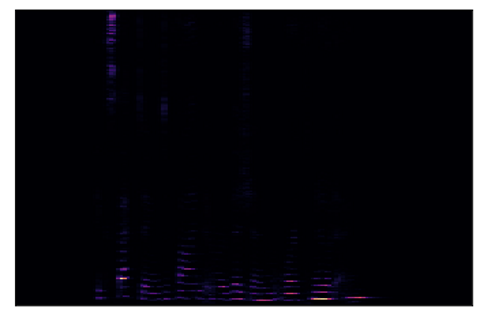


Figure 1.6

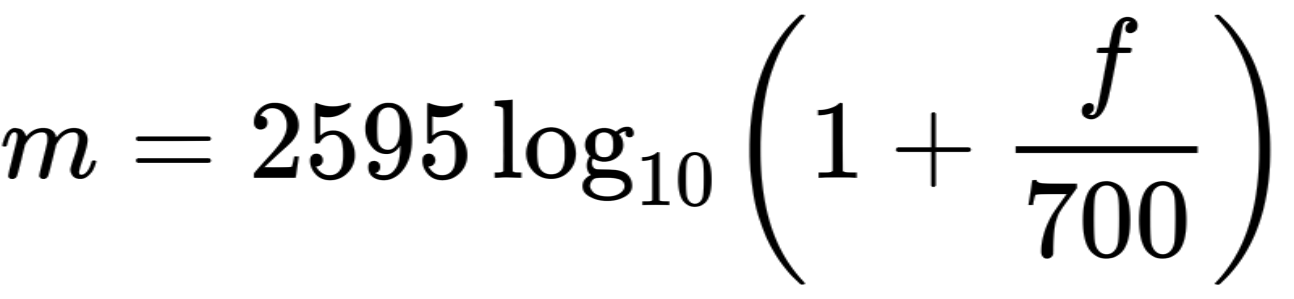
**1.9MELSCALE**The mel scale is a scale of pitches judged by listeners to be equal in distance one from another. The reference point between this scale and normal frequency measurement is defined by equating a 1000 Hz tone, 40 dB above the listener's threshold, with a pitch of 1000 mels. Below about 500 Hz the mel and hertz scales coincide; above that, larger and larger Intervals are judged by listeners to produce equal pitch increments.



Figure 1.7

**Formula**

A popular formula to convert f hertz into m mels is



**1.10 MEL Spectrogram**

Mel spectrogram is a spectrogram that is converted to a Mel scale. Then, what is the spectrogram and The Mel Scale? A spectrogram is a visualization of the frequency spectrum of a signal, where the frequency spectrum of a signal is the frequency range that is contained by the signal. The Mel scale mimics how the human ear works, with research showing humans don’t perceive frequencies on a linear scale. Humans are better at detecting differences at lower frequencies than at higher frequencies.

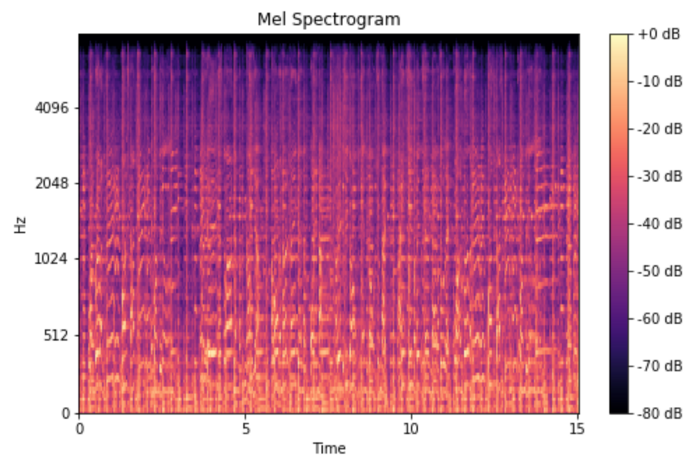


Figure 1.8

A mel spectrogram differs from a linearly scaled audio spectrogram in two ways:

1. A mel spectrogram logarithmically renders frequencies above a certain threshold (the corner frequency). For example, in the linearly scaled spectrogram, the vertical space between 1,000 and 2,000Hz is half of the vertical space between 2,000Hz and 4,000Hz. In the mel spectrogram, the space between those ranges is approximately the same. This scaling is analogous to human hearing, where we find it easier to distinguish between similar low frequency sounds than similar high frequency sounds.
2. A mel spectrogram computes its output by multiplying frequency-domain values by a filter bank.

**1.11 Histogram**

A histogram is a graphical representation that organizes a group of data points into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins.

Histograms are commonly used in statistics to demonstrate how many of a certain type of variable occurs within a specific range. For example, a census focused on the demography of a country may use a histogram to show how many people are between the ages of 0 - 10, 11 - 20, 21 - 30, 31 - 40, 41 - 50, etc.

**1.12 Histogram VS Spectrogram**

A histogram basically depicts an estimate of the probability distribution of some variable. To construct a histogram, the range of possible variable values gets divided into a series of intervals called bins. The bins must be adjacent to each other and are often (but necessarily) of equal width. Then a count of how many values fall into each interval determines the height of each bin such that the height is proportional to the number of cases in each bin. A histogram may also be normalized to display “relative” frequencies. It then shows the proportion of cases that fall into each of several categories, with the sum of the heights equaling one.

Spectrogram is a visual representation of the spectrum of frequencies found in a signal as they vary with time. Spectrograms of audio frequencies are sometimes called voiceprints or voicegrams. When the data is represented in a 3D plot the resulting depiction may be called a waterfall.

**1.13 Neural Network**

A neural network is a series of algorithms that endeavours to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature

A neural network works similarly to the human brain’s neural network. A “neuron” in a neural network is a mathematical function that collects and classifies information according to a specific architecture. The network bears a strong resemblance to statistical methods such as curve fitting and regression analysis.

**Architecture of Neural Network**

There are three main components: an input later, a processing layer, and an output layer. The inputs may be weighted based on various criteria. Within the processing layer, which is hidden from view, there are nodes and connections between these nodes, meant to be analogous to the neurons and synapses in an animal brain.

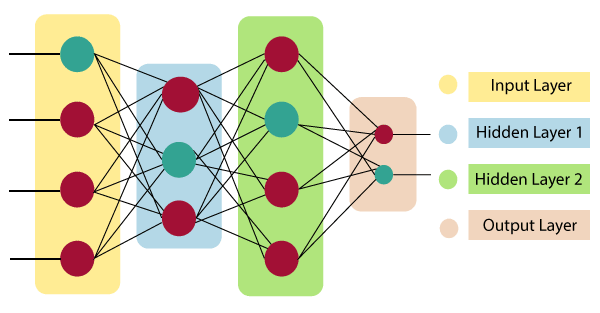


Figure 1.9

**Input Layer:**

As the name suggests, it accepts inputs in several different formats provided by the programmer.

**Hidden Layer:**

The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

**Output Layer:**

The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

**How do artificial neural networks work?**

Neural Network can be best represented as a weighted directed graph, where the artificial neurons form the nodes. The association between the neurons outputs and neuron inputs can be viewed as the directed edges with weights. The Artificial Neural Network receives the input signal from the external source in the form of a pattern and image in the form of a vector. These inputs are then mathematically assigned by the notations x(n) for every n number of inputs.

Afterward, each of the input is multiplied by its corresponding weights ( these weights are the details utilized by the artificial neural networks to solve a specific problem ). In general terms, these weights normally represent the strength of the interconnection between neurons inside the artificial neural network. All the weighted inputs are summarized inside the computing unit.

If the weighted sum is equal to zero, then bias is added to make the output non-zero or something else to scale up to the system's response. Bias has the same input, and weight equals to 1. Here the total of weighted inputs can be in the range of 0 to positive infinity. Here, to keep the response in the limits of the desired value, a certain maximum value is benchmarked, and the total of weighted inputs is passed through the activation function.

The activation function refers to the set of transfer functions used to achieve the desired output. There is a different kind of the activation function, but primarily either linear or non-linear sets of functions. Some of the commonly used sets of activation functions are the Binary, linear, and Tan hyperbolic sigmoidal activation functions.

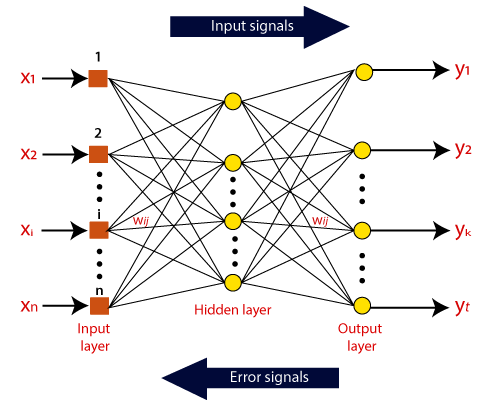


Figure 1.10

**Binary:**

In binary activation function, the output is either a one or a 0. Here, to accomplish this, there is a threshold value set up. If the net weighted input of neurons is more than 1, then the final output of the activation function is returned as one or else the output is returned as 0.

**Sigmoidal Hyperbolic:**

The Sigmoidal Hyperbola function is generally seen as an "S" shaped curve. Here the tan hyperbolic function is used to approximate output from the actual net input. The function is defined as:

F(x) = (1/1 + exp(-????x))

Where ???? is considered the Steepness parameter.

**1.13.1 CNN( Convolutional Neural Network)**

A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. It’s also known as a ConvNet. A convolutional neural network is used to detect and classify objects in an image.

CNN are very satisfactory at picking up on design in the input image, such as lines, gradients, circles, or even eyes and faces.CNN contains many convolutional layers assembled on top of each other, each one competent of recognizing more sophisticated shapes. Convolutional neural network, a convolution is a linear operation that involves the multiplication of a set of weights with the input, much like a traditional neural network. Given that the technique was designed for two-dimensional input, the multiplication is performed between an array of input data and a two-dimensional array of weights, called a filter or a kernel.

The filter is smaller than the input data and the type of multiplication applied between a filter-sized patch of the input and the filter is a dot product. A dot product is the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed, always resulting in a single value. Because it results in a single value, the operation is often referred to as the “scalar product“.

**1.13.2 MobileNetV2**

The MobileNetV2 architecture utilizes an inverted residual structure where the input and output of the residual blocks are thin bottleneck layers. It also uses lightweight convolutions to filter features in the expansion layer. Finally, it removes non-itineraries in the narrow layers. The overall architecture looks something like this.

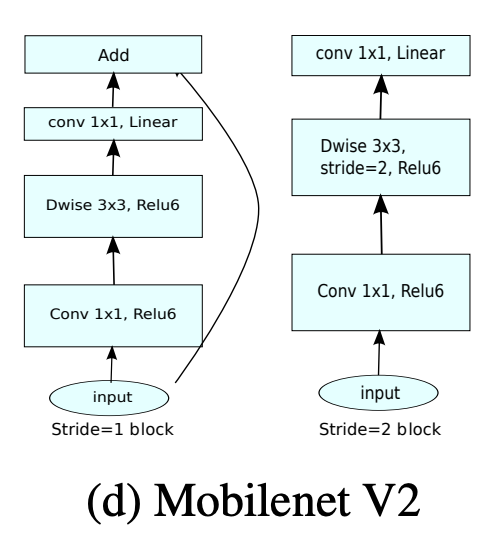


Figure 1.11

**CHAPTER 2**

**REQUIREMENT ANALYSIS**

* 1. **Proposed Work**

There are mainly four steps data pre-processing, feature extraction, classification, and recommendation system. The Feature extraction and data pre-processing is performed using Librosa python Library. The Librosa library performs both data preprocessing and feature extraction. The CNN is used for classification tasks.

* + 1. **Audio Sample**

The data from the GTZAN dataset is split into train and test in the ratio of 80-20 ratio. The GTZAN dataset is a collection of 9 music genres with 90 audio files each. All having a length of 30 seconds. It is the most used public dataset for evaluation in machine learning research for music genre recognition. The files are collected in 2000-2001 from a variety of sources including personal CD’s, radio microphone recordings in order to represent a variety of recording conditions.

The songs are splitting using a sliding window of 2.56s with 50% overlapping. The moving window will slide each song, then extract much length of the song and saved it into a list. Parallelly the label of that song also will be saved in another list. There are 9 classes of music genres such as blues, classical, country, rock, disco, reggae, hip-hop,metal and pop. The audio files having a similar size and similar ranges. The total number of songs in a dataset is 900 audio files.

* + 1. **Data Preprocessing**

The split\_convert function calls both to\_Mel-spectrogram and split songs functions. The split\_convert function is used to splitting the songs into small chunks. These small chunks are fed into the Librosa python library. The Librosa library is used to visualize the spectrograms. The Librosa library internally uses a Fast Fourier Transform (FFT) algorithm to analyze the audio signals in time domain and convert spectrograms in the frequency domain. The to\_Mel-spectrogram function is used to convert small chunks of songs into spectrograms and assign labels to small chunks. Before Feature extraction, feature scaling and hot encoding is performed. Feature scaling is the process of normalizing the data ranging between 0 and 1. The one-hot encoding is a method used to convert categorical data into integer data.

* + 1. **Feature Extraction**

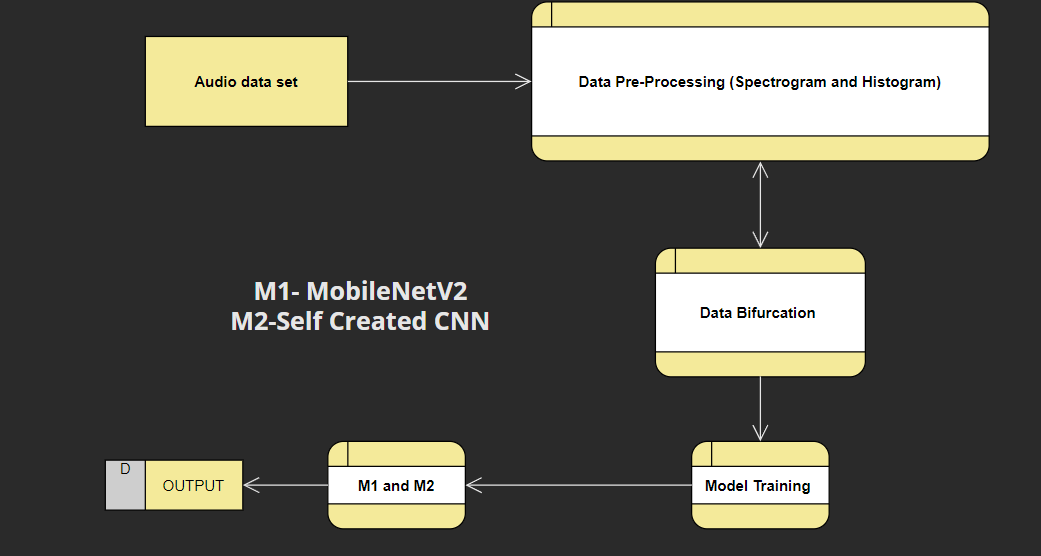
In feature extraction, extracting meaningful features from slices of spectrograms such as Mel-spectrograms, Specentroid, Spectral roll-off, Zero-crossing rate, Spectral bandwidth, and Chromo frequencies. Feature extraction valuable activity for analyzing and understanding the relations between songs and genres. Extracting the feature values from the slices of songs using Librosa python library. The Librosa python library performs feature extraction using different signal processing techniques. The mel\_spectrograms function not to convert spectrograms into images and only using the spectrogram features. These features are stored in a NumPy array(arr\_spec) with all the numerical values comprising the spectrogram. The arr\_genres are composed of categorical variables of genres uptown 0 to 9. After feature extraction, data augmentation is performed. The function GTZAN generator is created using the sequence method from TensorFlow. It belongs to the data generator classes. The advantage of using data augmentation is to increase the amount of training data. It reduces overfitting and gradually increases the accuracy of the model. In the augment function, horizontal flip and the random cutout are used to improve the size of the training data. Horizontal flip is used to flipping the spectrogram feature NumPy array and create a new input. In random cutout, a random NumPy array is generated and dB values are filled in the places to make it an actual spectrogram signal in the frequency domain.

**2.1.4 Classification**

Compile the model by giving the loss function and optimizer. The optimizer is used as adam optimizer. These will be the parameters for the neural network training process. ReduceLRonPlateau is used to reduce the learning rate when a matric has stopped improving. Batch size is mentioned which takes into a number of input samples at each epoch of training. More batch size can increase the accuracy as well as training time. Train the model for 200 epochs. The training is done purely in CPU using TensorFlow. After training, the model is saved using the function model.save function in Keras. For testing purposes, reloading the model. Then pass the test data into it and generate the model accuracy classification report and confusion matrix. Next stage, put the file want to test in the ’unknown’ folder. Then reuse the read\_data function to read the unknown music. This will extract the features from the music and convert it

into a format that is readable for the neural network. The unknown data is split into sub-samples. Each sub-samples is then passed into the neural network to predict and give the output label. From the output labels, implement a majority voting system that checks which genre has the most count and then classifies the music belonging to that genre. There are 30 slices of unknown audio that will pass into the model and predict the output. All the 30 predicted labels are added to a list. Use a concept of dictionary element in python to find out which label is the most predicted from among the 9 labels.

* 1. **Data Flow Diagram**



**CHAPTER 3**

**PROJECT DESIGN**

**3.1 Software Used**

1. **Anaconda-** Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, both of which are not free.

Anaconda distribution comes with over 250 packages automatically installed, and over 7,500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator, as a graphical alternative to the command-line interface

1. **Spider-** Spyder is a free and open source scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

**3.2 Important Libraries**

1. **Librosa-** Librosa is a Python package for music and audio analysis. Librosa is basically used when we work with audio data like in music generation(using LSTM’s), Automatic Speech Recognition.

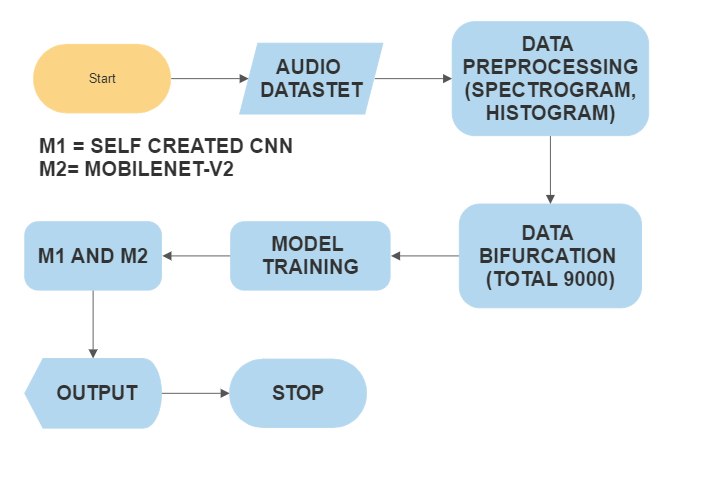
It provides the building blocks necessary to create the music information retrieval systems. Librosa helps to visualize the audio signals and also do the feature extractions in it using different signal processing techniques.

1. **NumPy -** NumPy (pronounced (NUM-py) or sometimes(NUM-pee)) is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors. NumPy is a NumFOCUS fiscally sponsored project.
2. **Tenser Flow-**TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.TensorFlow was developed by the Google Brain team for internal Google use in research and production.The initial version was released under the Apache License 2.0 in 2015.Google released the updated version of TensorFlow, named TensorFlow 2.0, in September 2019.TensorFlow can be used in a wide variety of programming languages, most notably Python, as well as Javascript, C++, and Java. This flexibility lends itself to a range of applications in many different sectors.
3. **Keras-** Keras is an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.Up until version 2.3, Keras supported multiple back ends, including TensorFlow, Microsoft Cognitive Toolkit, Theano, and Plaid Ml. As of version 2.4, only TensorFlow is supported. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. It was developed as part of the research effort of project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System),and its primary author and maintainer is François Chollet, a Google engineer. Chollet is also the author of the XCeption deep neural network model.
4. **Matplotlib**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged.SciPy makes use of Matplotlib.

Matplotlib was originally written by John D. Hunter. Since then it has an active development community and is distributed under a BSD-style license. Michael Droettboom was nominated as matplotlib's lead developer shortly before John Hunter's death in August 2012 and was further joined by Thomas Caswell.Matplotlib is a NumFOCUS fiscally sponsored project.

**3.3 Flow Chart**



**CHAPTER 4**

**RESULT AND OUTPUT**

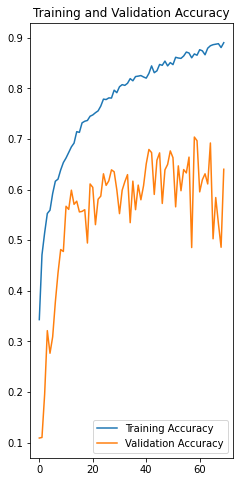
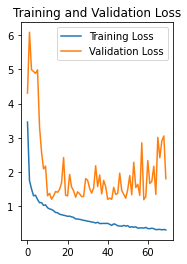
**4.1 Spectrogram**

|  |  |
| --- | --- |
| **Music Genre** | **Picture Of Spectrogram** |
| **BLUES** |  |
| **CLASSICAL** |  |
| **COUNTRY** |  |
| **DISCO** |  |
| **HIPHOP** |  |
| **METAL** |  |
| **POP** |  |
| **REGGAE** |  |
| **ROCK** |  |

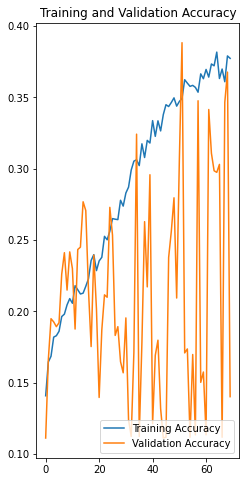
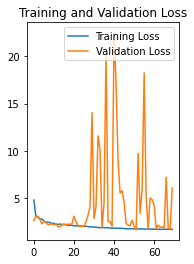
**4.2 Histograms**

|  |  |
| --- | --- |
| **Music Genre** | **Picture of Histograms** |
| **BLUES** |  |
| **CLASSICAL** |  |
| **COUNTRY** |  |
| **DISCO** |  |
| **HIPHOP** |  |
| **METAL** |  |
| **POP** |  |
| **REGGAE** |  |
| **ROCK** |  |

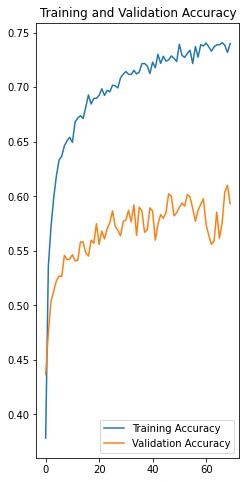
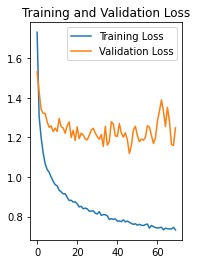
**4.3 Accuracy and Loss Graph for Spectrogram using CNN**

****

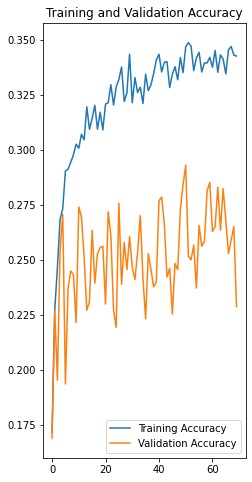
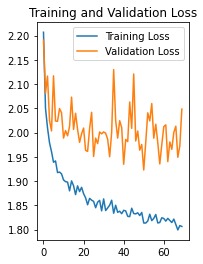
**4.4 Accuracy and Loss Graph for Histogram using CNN**

****

**4.5 Accuracy and Loss Graph for Spectrogram using MobileNetV2**

** **

**4.6 Accuracy and Loss Graph for Histogram using MobileNetV2**

** **

**CHAPTER 5**

**CONCLUSION AND FUTURE WORK**

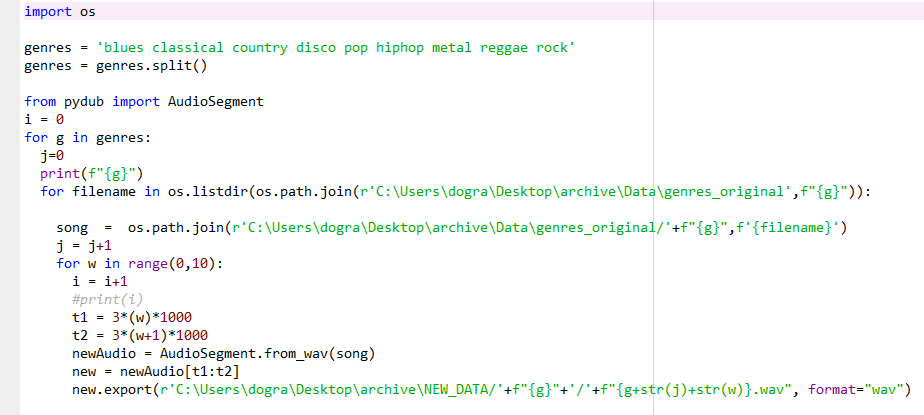
**Conclusion-** Under this experiment, two different CNN models such as MobileNetV2, and a self-created model are used and both of them were trained on the same dataset and observed that how they work when performed on the same dataset. Through this work, we concluded that self-created model perform better on the same dataset as the accuracy for Spectrogram is 90 percent and for Histogram is 37 percent. Refer to the accuracy graph for reference. Thus, it is concluded that for Spectrogram both the models have given good training accuracy score whereas for Histogram none have reached over 40 percent. This work can be further extended to get better objectives by applying some image pre-processing techniques before training the model to get better accuracy. Increasing the value of RGB images remains an important factor that needs to be kept in mind.

**Future Work-** Future work for this project would be a recommendation system. This recommendation system will tell the genre of the upcoming music and would also recommend us to listen which song. And for the recommendation part we can use Linear Regression for that which can predict the genre of the upcoming music.

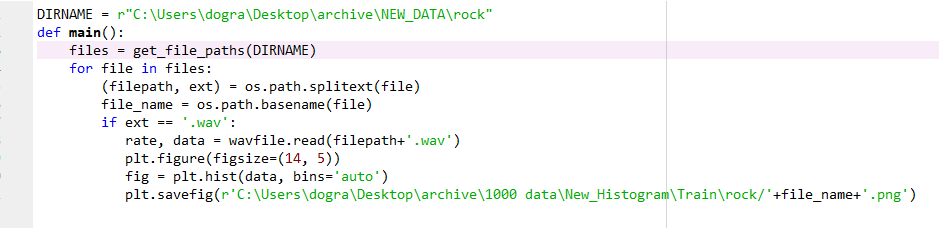
**APPENDIX**

**(CODE)**

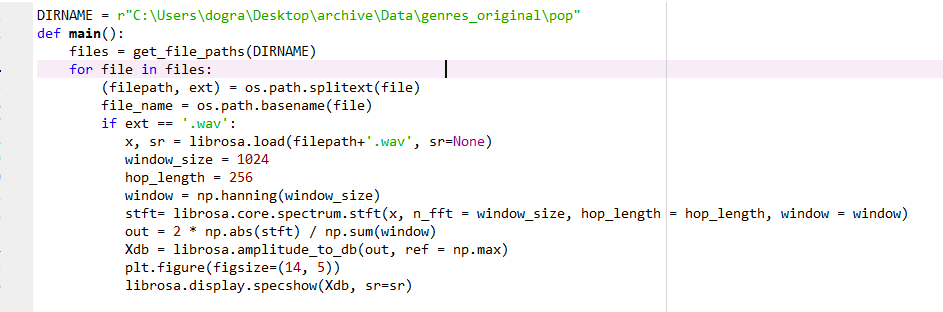
**Dividing Audio into 10 parts**

****

**Code for Histogram**

****

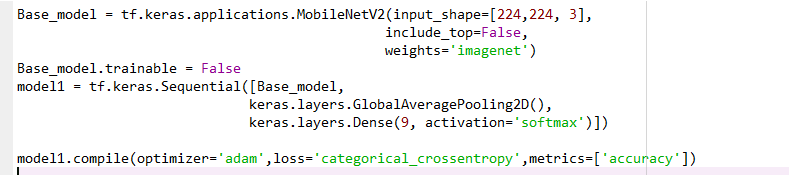
**Code for Spectrogram**

****

**Self Created CNN**

****

**MobileNet-V2**

****

**REFERENCE**

[1] Deep learning based music genre classification using spectrogram

Athulya K Ma, Sindhu Sb PG Student, Dept. of computer science and Engineering, NSS College of Engineering Palakkad, Kerala

bProfessor

[2] Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua

Tao, and Lian-Hong Cai. 2002. Music type classi-fication by spectral contrast feature. In Multimedia and Expo, 2002. ICME’02. Proceedings. 2002 IEEE International Conference on. IEEE, volume 1, pages 113–116.