Indian Classical Music Analysis and Understanding

Polishetty Akshay, polishettyakshay@gmail.com

ABSTRACT

Indian classical music is the music of the Indian subcontinent and is one of the oldest kinds of music. Although Indian music has progressed significantly, the fundamental components appear to remain the same as they were two thousand years ago. The raga is at the center of Indian classical music, and it is the language of the spirit. Novel Indian music, on the other hand, has been heavily impacted by Western music notation, which is based on equal mean tones, temperament scales, and the universal priority of harmony. Indian and Western music differs not just in terms of culture, but also in terms of fundamental structures, scales, and tuning. New musical innovations are burgeoning in the current decade with an incredible number of fusions of Western Classical music and Indian Classical music. While fusions are a way of celebrating two different art forms, is the originality of the classical forms fading away? This project aims at studying the classification of Indian Classical Music and Western Classical music. A simple feed forward neural network was created to classify the audio data. The audio classification is not limited to the style of music. The model also classifies the audios according to the instruments and the different ensembles at play. A testing accuracy of 75.00 percent and a training accuracy of 97.9 percent was achieved through the model. This project is an attempt to encourage musical fusions and to help create awareness among the community about the ancient and widely practiced art of Carnatic Music.

INTRODUCTION

Music has always been more than just sound—it's a universal language that goes beyond words and borders. Since ancient times, music has played an important role in shaping cultures and bringing people together. In the beginning, it was often seen through the lens of just one culture, but over time, people started to appreciate music from different parts of the world. This shift gave rise to a more multicultural view of music, where traditions from around the globe are celebrated and shared [1].

Indian classical music, for example, has a deep history that goes back over 6,000 years. It started with the Vedic scriptures, where chants gradually evolved into a complex system of musical notes and rhythms. This form of music is deeply connected to nature—many of its melodies, called *ragas*, are inspired by the changing seasons or specific times of the day. Its rhythmic patterns, known as *taals*, are also highly structured and symbolic [2].

Carnatic music, which is the classical music of South India, gets its name from the Sanskrit term *Karnātaka Sangītam*, meaning “traditional” or “codified” music [3]. It developed in the southern states of Tamil Nadu, Kerala, Andhra Pradesh, and Karnataka—all regions rich in Dravidian heritage. The foundations of modern Carnatic music were laid by Purandara Dasa in the 15th century, who systematized the teaching methods [4]. Later, Venkatamakhin (also known as Venkat Mukhi Swami) made a major contribution by creating the *Melakarta* system, which is still used today to classify ragas [5].

On the other hand, Western music reflects the cultures of Europe, the Americas, and other places influenced by European settlers. Its roots can be traced back to the 19th century. One thing that sets it apart from Indian classical music is its use of major and minor scales and equal temperament tuning. Western music typically has seven main notes and five variations, arranged in increasing pitch to form what's known as an octave [6].

METHODOLOGY

This section focuses on the core aspect of the experiment—data collection and usage. Due to the lack of a standard dataset that includes both Carnatic classical and Western classical music along with the ensembles and instruments used, a custom dataset was developed. For Carnatic music, data was sourced from Saraga's Indian Art Music dataset (version 1.5 - Carnatic) [7]. From this extensive collection, nine songs were randomly selected from four instrument categories: violin, mridangam (left-hand), ghatam, and vocals. These original recordings varied in length, ranging from 30 seconds to 50 minutes, so each was trimmed to 10–15-second clips. The trimming process was carried out randomly using iMovie and Google's random number generator (RNG), ensuring uniformity and minimizing selection bias. Similarly, Western classical music samples were collected from the MusicNet dataset (version 1.0) [8], focusing on six specific ensemble types: Solo Piano, String Quartet, Solo Violin, Wind Quintet, Solo Cello, and Accompanied Violin. Nine samples from each category were selected and shortened to 10–15 seconds using the same randomization method. All audio files were saved in .wav format with an average sample rate of 24 bits per second or 48 kHz. After organizing the audio clips, a .csv file was created containing metadata such as file name, salience, folder number, class ID, and class label. Structuring the dataset in this way made it easier for the classification model to read the files efficiently using Pandas and helped reduce the need for hardcoded file paths in the code.

CODE AND ML MODEL

The code for this project is built using a neural network with Keras and TensorFlow as the backend [9]. It starts by importing all the necessary libraries such as NumPy, pandas, librosa, matplotlib, and the relevant Keras modules needed to build the model. Once the environment is set up, the dataset is loaded by reading a .csv file using pandas. To give a quick visual overview, one audio file from each class is extracted and its waveform is displayed.

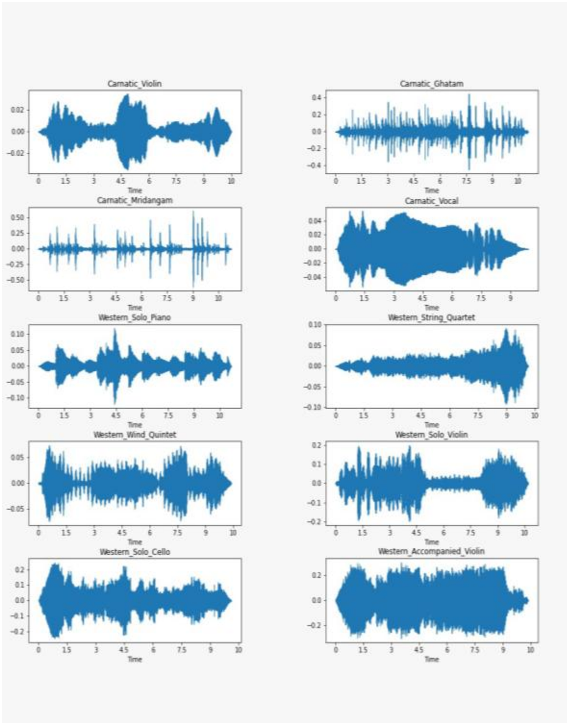
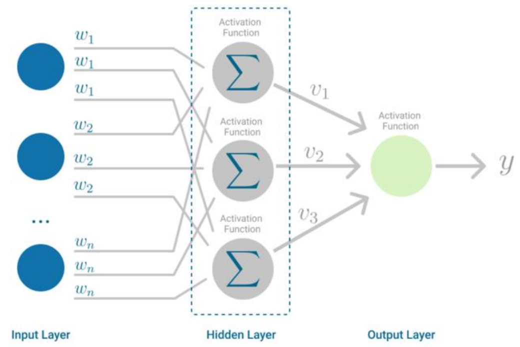
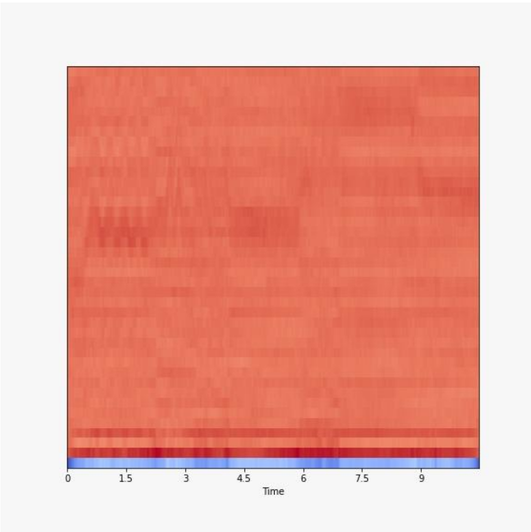


Figure 1 shows the waveforms of 10 audio files from each music class. These waveforms visually represent the sound, with amplitude on the Y-axis and time on the X-axis. The first column (top to bottom) includes waveforms from: Carnatic Violin, Carnatic Mridangam, Solo Piano, Western Wind Quintet, and Western Solo Cello. In the second column (top to bottom), you'll find: Carnatic Ghatam, Carnatic Vocal, Western String Quartet, Western Solo Violin, and Western Accompanied Violin.

After that, feature extraction begins using librosa and scipy.io. The librosa.load function plays a key role here—it not only standardizes the audio by converting the sample rate to 22.05 kHz but also ensures the audio is in mono format instead of stereo. One of the most important steps in this phase is extracting features in the form of MFCCs (Mel

Frequency Cepstral Coefficients), which are commonly used in audio processing. These coefficients capture the shape of the audio's spectral envelope and typically include 10–20 key features. In this case, 40 MFCCs were extracted from each audio clip, resulting in a 40-element array that represents the audio's characteristics.






```
array([-5.41517151e+02,  1.03288857e+02, -2.09685421e+01,  4.51897926e+01,
       -7.31762743e+00, -1.06241798e+01, -1.04500074e+01, -9.71995735e+00,
       -3.11487412e+00, -1.09566174e+01, -1.54824762e+01, -4.04470110e+00,
       -7.72584963e+00, -5.83546305e+00, -1.89279795e+00,  4.21685457e+00,
       -7.58998919e+00,  5.25713253e+00, -7.58154809e-01, -1.27665873e+01,
       -2.89532959e-01, -1.49172628e+00,  1.10488358e+01,  1.37288427e+01,
        1.06224051e+01,  1.29149094e+01,  3.51227188e+00,  8.99843121e+00,
        1.01063786e+01,  9.53975391e+00,  4.43502776e-02, -1.01534195e+01,
       -7.75432301e+00, -3.96182323e+00,  7.05239153e+00,  3.21995211e+00,
       -2.52421021e-01, -1.84600562e-01, -1.97548246e+00, -5.05416965e+00],
      dtype=float32)
```

Before training the model, the dataset was split randomly into training and testing sets using the train_test_split function—allocating 50 samples for training and 40 for testing out of the 90 total audio files. With the data ready, a simple feedforward neural network was built using Keras in a sequential fashion. This model falls under the category of Multi-layer Perceptrons (MLPs), which are designed to handle nonlinear relationships between input and output. The architecture includes layers such as Dense, Dropout, and Activation, all working together to learn and classify the audio features effectively.

Figure 4 illustrates the MLP (Multi-Layer Perceptron) model, which uses backpropagation to reduce the cost function and improve its performance [10].

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_4 (Dense)	(None, 256)	10496
activation_4 (Activation)	(None, 256)	0
dropout_3 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 256)	65792
activation_5 (Activation)	(None, 256)	0
dropout_4 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 10)	2570
activation_6 (Activation)	(None, 10)	0

Total params: 78,858
Trainable params: 78,858
Non-trainable params: 0

Figure 5 presents a summary of the model's performance. Initially, the model started with a pretraining accuracy of 15.00%. After training for 134 epochs, it achieved a training accuracy of 97.90% and a testing accuracy of 75.00%.

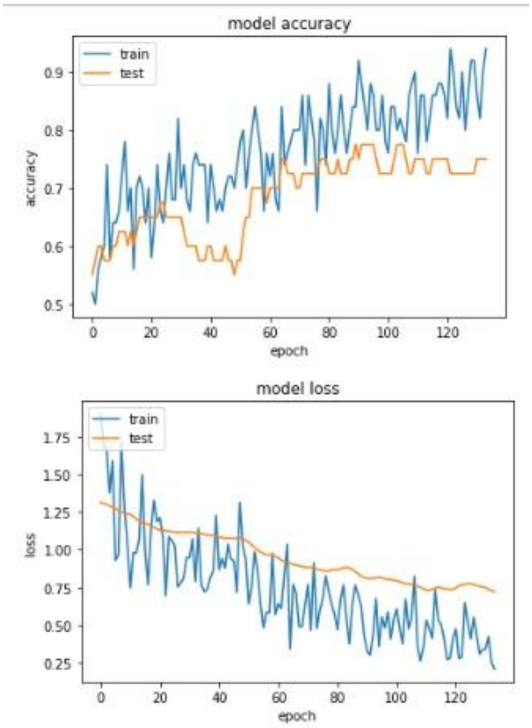


Figure 6 shows the model's accuracy and loss over the training process, giving insight into how well the model learned and how its performance improved over time.

ANALYSIS THROUGH COMET.ML

To gain a deeper understanding of the model's performance—specifically accuracy versus loss—**Comet.ml** [11] was integrated into the workflow. It logged all relevant data during the code's compilation and execution. The following graphs present a detailed breakdown of this analysis.

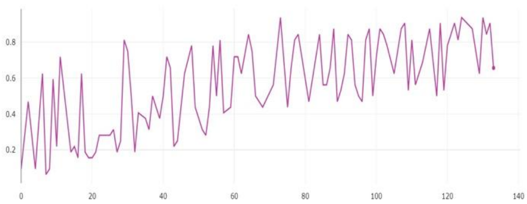


Figure 7 displays the training accuracy of the model over time, with epochs on the X-axis and accuracy (in %) on the Y-axis.

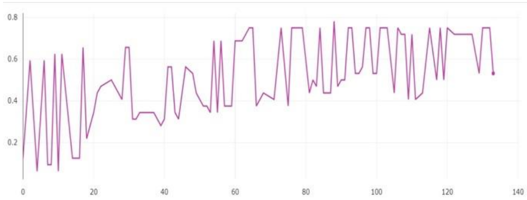


Figure 8 shows the testing accuracy across different epochs, with the X-axis representing the number of epochs and the Y-axis showing accuracy in percentage.

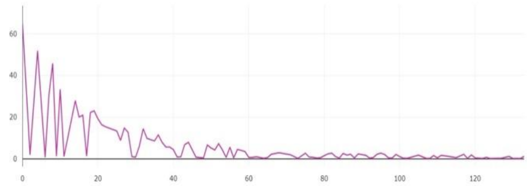


Figure 9 illustrates the training loss over the course of training, with epochs on the X-axis and loss (in %) on the Y-axis.

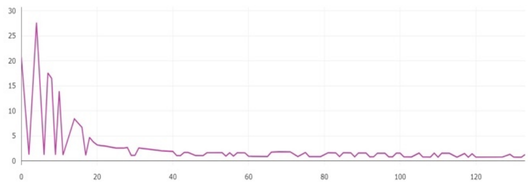


Figure 10 shows how the testing loss changed over the epochs, with the X-axis representing the number of epochs and the Y-axis indicating the loss percentage.

FUTURE WORKS

This project is still in its early stages and has a long way to go to fully achieve its intended goals. However, there's plenty of room for future development. The work can be expanded in several exciting directions—such as examining whether the authenticity of traditional classical music is preserved in emerging genres and fusion music, conducting comparative analyses of fusion styles, exploring mathematical combinations of different genres, and identifying the distinct genres present within a fusion composition.

ACKOWLEDGEMENTS

I would like to express my sincere gratitude to Polygence for giving me the opportunity to work on a project aligned with my career interests. A special thank you goes to my mentor, Mohamed Eltaeb, for his guidance and support throughout my research. I’m also deeply thankful to Dr. Ramaprasad Poojary for his insightful and constructive suggestions, which were incredibly valuable to the progress of my work.

REFERENCES

[1] *The Cultural Connection.* UUA.org. (2022). Link

[2] *An Introduction to Indian Classical Music.* Making Music. (2022).

[3] *Carnatic Music - Carnatic Classical Music - Indian.* Cultural India. Link

[4] *Indian Music - Genres, History & Evolution.* Cultural India.

[5] *Carnatic Vocals Classes.* Saanwee Performing Arts | United States.

[6] Khillar, S. (2022). *Difference Between Indian Music Notes and Western Music Notes.* Difference Between.

[7] Bozkurt, B., Srinivasamurthy, A., Gulati, S., & Serra, X. (2018). *Saraga: Research Datasets of Indian Art Music (1.5)* [Data set]. Zenodo.

[8] John Thickstun, Zaid Harchaoui, & Sham M. Kakade. (2016). *MusicNet (1.0)* [Data set]. Zenodo.

[9] Mendels, Gideon. (November 18, 2019). *How to Apply Machine Learning and Deep Learning Methods to Audio Analysis.* Medium, Towards Data Science.

[10] *Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis.* Medium. (2022).

[11] *Comet | Build Better Models Faster.* Comet. (2022).