A

ARTIFICIAL INTELLIGENCE PROJECT REPORT

on

Harmony Hub

Submitted by:

Khushi (220584)

Payal Dabas (220314)

Sai Tejaswi Woonna (220499)

Abhishek Sharma (220558)

under mentorship of

Dr. Atul Mishra
(Assistant Professor)

Department of Computer Science Engineering
School of Engineering and Technology
BML MUNJAL UNIVERSITY, GURUGRAM (INDIA)

November 2024

Candidate's Declaration

I hereby certify that the work on the project entitled, "**Harmony Hub – Instrument Tutoring System**", in partial fulfillment of requirements for the award of Degree of **Bachelor of Technology** in School of Engineering and Technology at BML Munjal University, having University Roll No. 220584, 220314, 220499, 220558, is an authentic record of my own work carried out during a period from August 2024 to November 2024 under the supervision of DR. ATUL MISHRA.

(Khushi)

(Payal Dabas)

(Sai Tejaswi Woonna)

(Abhishek Sharma)

Supervisor's Declaration

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Faculty Supervisor Name: Dr. Atul Mishra

Signature:

ABSTRACT

Harmony Hub project is an intelligent instrument tutoring system, which incorporates audio classification based on deep learning algorithm. The final data set will be created by using the training and testing set with audio files and accordingly with the different types of instruments. In this project the Mel Frequency Cepstral Coefficients are used in feature extraction and pre-processing. The trained CNNs and LSTM networks are then feed into processed data, which forms good input for an audio signal, as it captures the spatial and temporal characteristics of a signal. The main measure, by which the performance of the system is assessed, is accuracy. The integration of these contemporary methods in audio signal processing with recent advancements in machine learning form the strong base for automatic detection of instruments and thus creates a paradigm shift for high level personalized music learning applications.

ACKNOWLEDGEMENT

I am highly grateful to **Dr. Atul Mishra, Assistant Professor**, BML Munjal University, Gurugram, for providing supervision to carry out the seminar/case study from August-November 2024.

He has provided great help in carrying out my work and is acknowledged with reverential thanks. Without wise counsel and able guidance, it would have been impossible to complete the training in this manner.

I would like to express thanks profusely to thank **Dr. Atul Mishra**, for stimulating me from time to time. I would also like to thank the entire team at BML Munjal University. I would also thank my friends who devoted their valuable time and helped me in all possible ways toward successful completion.

(Khushi)

(Payal Dabas)

(Sai Tejaswi Woonna)

(Abhishek Sharma)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

The idea of the Harmony Hub project is to develop an algorithm in the sphere of instrument tutoring using the machine learning approach to classify the acquired signal from a variety of musical instruments. Feature extraction of speech signal is done by Mel Frequency Cepstral Coefficients (MFCC) while the signal classification us done with Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. It will even be capable to identify the instruments being played and offer prompt personalized feedback to the learner. This would make music education even more accessible and scalable while empowering an inventive way that uses Audio Signal Processing with Machine Learning for a new level of learning experience.

1.1 Problem Statement

Learning of musical instruments and traditional music learning in general has always been done according to what is referred to as face-to-face learning, which is both time consuming and very expensive and may, as such, not be easily available to almost anyone. Furthermore, real-time responses and personal training are the reasons with which new entrants cannot improve the skill proficiently. Consequently, there is a demand for an automatically working system that can recognize musical instruments with help of audio analysis and offer an interactively response in learning music.

1.2 Objective

The principle goal of the Harmony Hub is to create a fully intelligent system that can classify different audio signals taken from musicians' distinct instruments. When applying machine learning, the system's goals include the recognition of the different instruments as well as giving feedback in real time in order to mimic an intelligent tutoring system. It will be based on deep learning model like CNN, LSTM networks using MFCC for features along with such kind of concepts to make a highly efficient and accurate instrument recognition system.

1.3 Motivation

This idea stems from the increasing demand and desire for non-traditional, affordable and more convenient approaches to learning. Clearly, there is great opportunity for applying machine learning and audio signal processing where emerging technologies will change how children and adults learn and teach music. Here, at Harmony Hub it can let learning music much more engaging and effective for the learner through automation of instrument identification or feedback irrespective of location or financial capabilities of the learner.

1.4 Significance

However, the general impact of this project here lies in the modality of transforming the delivery of the subject of music education. Two areas that could make Harmony Hub very useful for self-learners and educators is its ability to automatically recognize instruments to provide feedback with proper outputs and it would enable the learners to exercise them as assisted by an intelligent and always-present system; hence, this would complement the conventional human music tuition and supply it with sophisticated technology. As outcomes, it may generate efficient methods that will help people of different cultures and age to learn music online.

1.5 Challenges

This problem presents a high level of problem complexity from the audio signal processing and feature extraction perspective. The frequencies and sounds from different instruments are similar and sometime overlap with each other and therefore makes the classification of the data sets slightly challenging. However it should also be designed to handle changes in environment, noise, and differences in performances. Another issue is to fine-tune the deep learning models to enable the system provide fast and accurate classification to the learners during learning.

1.6 Novelty Proposed

The idea that has not been proposed before is the usage of ultra modern machine learning models such as CNNs and LSTMs along with traditional methods like MFCC for the proper categorization of instruments as well as instrument tutoring. The new possibility of applying deep learning-based approaches within the framework of the integration of spatial and temporal data analysis in FMA will create the conditions for enhancing the possibility of more precise instruments classification. The redirecting role in reforming the structure of this project aims at developing an interactive, automatically tutored platform for giving individualized feedbacks regarding learning goals as well as to improve the usage of the information and educational production and make musical education more accessible to a wider populace.

Chapter 2

Literature Review

This paper consists of the findings of the literature review on intelligent music coaching applications based on artificial intelligence. The following table categorizes the studies according to the focus areas by indicating the knowledge contributions, research methods, and findings. This structured framework gives a good understanding on the current development on audio processing, integrating AI into music learning and teaching, and interactive learning.

Category	Paper Title	Author	Key Contribution	Outcomes
Audio Processing	Fundamentals of Music Processing [1]	Müller, M. (2015)	The basic methods of the extracting the relevant features of the audio data and for pitch detection and rhythm analysis.	Inaugurated techniques for the description of musical components
	Deep Learning for Music Information Retrieval [2]	Bello, J. P., et al. (2018)	Better algorithms for deep learning for music as a combination of better identification of instruments and better pitch detection.	Availability of more precise information on Music Information Retrieval.
Instrument Recognition	The IRMAS Dataset [3]	Bosch, J., et al. (2016)	Broad dataset and primary measures for	Since the recognition methods were applied to a

			instrument detection.	standardized dataset, comparing the results was easier.
	Deep Convolutional Neural Networks for Predominant Instrument Recognition [4]	Han, Y., et al. (2017)	New architectures of CNN that have been designed reach 91% in the classification of instrument.	The main experimental result is the policy that achieves instrument recognition with a 91% accuracy.

Table 1: Literature review

- **Research Gap 1:** Long-term Effectiveness of AI Tutoring

  Key observation: Few longitudinal studies and international research findings on AI's effects on music instruction.

  Outcome: Suggested integration with traditional teaching methods to address gaps

- **Research Gap 2:** Technical Limitations

  Key observation: Some difficulties of polyphonic analysis; working with complex structure

  Outcome: Suggested improvements in the computational models

Chapter 3
Exploratory Data Analysis

3.1 Dataset

Therefore, the audio dataset used for the Harmony Hub project is the one consisting of various instruments: for instance violin, piano, guitar, flute, trumpet etc. The audio records of these instruments are split into folders being used for training and testing purposes and for each of these folders audio samples are provided by instrument category. This dataset is organized in a way that for each the text transcripts, there are corresponding audio files. Instrument will be correctly labeled and organized according to the type, for it to be effectively used for classification tasks.

Music samples for this research work were collected freely from audio databases or identified and downloaded. The coefficient and variance files define each of the classes with their instruments' category. Audio files were gathered from various sources to reduce sample bias and achieve data variation. Audio files are processed automatically through an automated pipeline involving several stages:

1. **Feature extraction:** For the retrieved audio, mel frequency cepstral coefficients (MFCC) are extracted from the sound, which helps in achieving a compact frequency domain representation of the sound.
2. **Data Augmentation:** The pitch shifting, time stretching, and noise addition are employed to augment the dataset and thus improve the model.

3.2 Exploratory Data Analysis

During the EDA phase, we check the distribution of instruments in the dataset in order to validate if it is balanced over categories. Pie charts and bar graphs assist in the representation of the number of samples per instrument. We also create spectrograms and plot the MFCC so that we are able to learn more about the audio data samples that we have collected. The plots convey the peculiarities of each sound from an instrument and so offers a clearer understanding. In other words, the features of reigns, which has been derived for the Violin would be are quite different from the one for the piano mainly because of the differences in the timbre and the tone and these differences are represented visually in order to inform the feature engineering process.

Feature extraction is one such crucial techniques employed to transform audio single channel data into usable data for machine learning applications so that they gather useful data for use in the machine learning models. Further to this, once the MFCC features are computed, additional feature extraction processes may be applied on them in cases where the application requires more elaborate discriminating factors. Evaluation is made to develop the connections between the coefficients and the associated instrument types. Interdependency is evident in the MFCCs for each of the analyzed audio files as follow tool, while correlation analysis is done to analyse these dependencies. This is especially ideal for use in the determination of feature selection when trying to segment methods such as the piano and violin is found. Other than MFCC, there are zero crossing rate, spectral centroid and to improve the feature set possibly, the behavior of roll-off in the time domain is investigated.

Chapter 4

Methodology

4.1 Problem Statement

Since most traditional instrumental lessons are relies on the functionality as well as results achieved are expensive, time-consuming and beyond reach for broad spectrums of students. Additionally, the current systems that support music learning contains significant deficits of individual feedback and real-time tutorial. Teachers which are crucial to the extent of helping the learners to be able to improve on what was taught successfully, the problem is that there is no widely-applicable, self-teaching system of music instruments which can offer instant and tailored feedback. Lacking such a system, many learners end up with conventional approaches that would slow down development and participation.

Relevance to AI and Real-World Applications: This is a problem related to the Artificial intelligence paradigm more specifically within the scope of machine learning and audio processing. It could be employed to broaden and expand the concept of music education by using artificial intelligence. The technology needed to develop systems that could identify the musical instruments and give feedbacks. Audio classification may receive real time results on instrument performance based on applications of AI such as CNN and LSTM networks which are being used in the project. In practical use such a system would be used by music schools, social networks, and budding musicians for more learning experiences. Hence, the incorporation of the AI in music education becomes kind of potential openings towards the more general tutoring.

4.2 State Space Search

4.2.1 State Space

- **States:** In this problem the factors that define a state are as follows:

1. **Audio Input:** The recordings of the signal that was obtained from the sound produced by the musical instrument in play.
2. **Extracted Features:** The features extracted from the raw audio include those that define the characteristics of the sound present in the microphones of the MFCC.
3. **Instrument Label:** The musical instruments class which the system is attempting to recognize, it could be violin, piano, guitar and so on.

Every state changes as system handle the audio and move to the next state. The three stages involve in the process are feature extraction, classification and the feedback.

- **Initial state:** This is represented by an audio file from the dataset which has not been classified or analyzed in this current stage. There were no extracted features and no information about which instrument was linked to the audio was available. Raw, unprocessed audio data was fed into the system.

- **Goal State:** The goal state is the correctly labelled instrument in which the system has classified the instrument that was played given all the features extracted from the audio file. In such a state, the system will give information to the learner regarding the performed instrument and if necessary, adjustments on the methods being used will be advised.
- **Possible Actions:** Possible actions in this state space search are the steps performed in order to prepare the raw audio and classify it:

1. **Preprocessing:** Record the same text and save it into another program in the simplest format, for instance, WAV, to avoid conversions or resampling if the sound is noisy.

2. **Feature Extraction:** It is recommended to extract features like MFCC for example in order to represent the audio data.
3. **Classification:** Apply the CNNs and LSTMs models in order to classify the instrument based on the selected information.
4. **Feedback Generation:** Provide feedback for the classification result that includes recommendation or the performance accuracy score.

4.2.2 Search Strategy

- **Description of the Chosen Algorithm:** Indeed, the problem of classifying the musical instruments with reference to audio data is best captured by the state space search wherein the system moves across potential features and arrives at the right instrument. In this regard, the search algorithm maps to a variant of best first search: a decision is incrementally refined based on learned patterns in the data. Instead of performing the search for each state and manually rank them, computational models, in our case, CNN-LSTM, mimic the process. Here, reminiscing CNN-LSTM architecture as a component of heuristic-based search, it is possible to state that the architecture optimizes on all levels and enhances both feature extraction and classification functions. However, there are mathematical approaches such as, during the feature selection or extraction, gradient-based optimization during training too, is implied. This optimization could therefore be viewed as an iterative, stochastic search of the parameter space in order to minimise classification error.

- **Justification and Implementation:** CNN-LSTM approach really serves the purpose of using the heuristic search on the higher feature space of the audio data. Spatial search tools are CNNs, that can recognize patterns in spectrogram like data of audio while the sequential traversal of sound data for instrument classification is done by LSTMs. This makes the combination fit nicely for classification tasks involving audio, besides spatial as well as temporal characteristics of the sound, it uses. Raw audio was preprocessed for extraction of MFCCs, which is a compressed feature that describes sound. Specifically, the CNN layers are used to extract spatial features and hierarchically refine spatial patterns, the LSTM layers are used after those for temporal patterns. The whole system is parameterized by such elements that update through gradient-based methods such as the Adam and iteratively approach the minimum value of the given classification error. This learning process enables the model to learn most relevant features and paths through the feature space which assists in successful classification of instruments as well as provision of real time feedback to learners.

4.3 Knowledge Representation

- **Representation Technique:** As part of the Harmony Hub Project, we apply an audio classification method that divides audio sounds into meaningful parts. In particular, audio signals are represented by Mel Frequency Cepstral Coefficients (MFCCs), which illustrate the spectral and temporal properties of sounds. Features provide a compact and easy way to encode the various sounds of different instruments by lowering the raw audio data without losing the important information needed for classification.

- **Implementation Details:**

1. **Audio Preprocessing:** All audio files are re-sampled to a common sample rate. To some degree, noise has been removed from the data in terms of two first methods.
2. **MFCC extraction:** It is first necessary to explain how MFCC features are calculated for every audio sample. First, the audio signal is segmented in time to form frames. Then, a fourier transform is performed to detect the frequency components of those frames which are transformed to the Mel scale to conform to the way humans perceive sound.

3. **Feature Normalization:** The MFCCs obtained are standardized in preparation for the input data. There is uniformity during the training process and this helps promote faster convergence of the machine learning model under consideration.
4. **Integration with CNN-LSTM:** As the MFCC features are transferred as an input to the CNN-LSTM model where the CNN learns the spatial features and the LSTM learns the spatial relationship. In order to get the precise class name of the instrument, it has to be properly classified.

- **Appropriateness and Justification:** The MFCC incorporates the most basic parameters because it is important to consider both perceptual and acoustical attributes of the sound as well as to make the representation needed for the separation. Here, it is important to note that this representation dilutes the input data's dimensionality which is beneficial for the computational aspect of machine learning models. In general, combination of MFCCs and the CNN-LSTM architecture provides opportunities to make use of spatial and the temporal-scale of the sound. The method is appropriate to this project because its primary goal is to identify a musical instrument with the hope that the user would be offered a response to it as soon as it is done. This method is appropriate to the project due to the fact that it is a close-up representation and it also requires the treatment of a complex audio signal.

4.4 Intelligent System Design

The system is created in the form of modular AI-driven architecture for instrument classification and feedback. The following are the layers included in the developed system:

The system accepts raw audio files uploaded by the user or in a dataset.

- **Preprocessing Layer:** Converts audio into a uniform format, removes noise, and extracts MFCC features from it.

- **The Model Layer:** The CNN-LSTM Hybrid model is applied for instrument classification. The Feedback Layer gives the user with classified results along with personalized feedback.
- **Interface Layer:** A user-friendly frontend for uploading audio files and receiving feedback. A backend API shall be supporting the layer for model inference and data processing.

Components and Functionalities:

The data processing pipeline:

a. audio waves are down-sampled and denoisied to standardize audio files.

b. And for input to a classifier, MFCCs are extracted as features.

The classification Engine:

a. CNN-LSTM Model:

- The encoders of the CNN layers extract spatial patterns of the sounds which are contained in the audio features.
- Sequential patterns in sounds which are crucial in the recognition process of sounds are captured by the LSTM layers in the model.

b. Finally, the model gives the final outputs which is the determined label of the instrument. Feedback Mechanism: Establish measures of effectiveness and formulates useful measures to increase the reliability of the users' classification.

User Interface: It enables the user to upload audio files and show the instantaneous classification and feedback results as also viewed in the interface.

Innovations:

- **Hybrid Model Integration:** It combines CNN and LSTM so that spatial and temporal audio features are employed effectively leading to better classification.
- **Real-Time Feedback:** It makes it possible to give immediate feedback to users, thus enhancing the learning processes.
- **Scalability:** Integrated modules also make it easy to up-scale with more instruments or enhanced features such as performance evaluation.
- **Human-AI Collaboration:** Combines techniques powered by AI with interactions aimed at users making it possible for the tutoring tool to be personal and easily accessible.

4.5 Constraint Satisfaction Problem

Variables:

- In the research, each audio file is viewed as a variable.
- Some features such as MFCC coefficients, pitch, timbre among others have been defined as sub-variables that are important in decision making.
- The classification output is also a variable whereby the system has to predict the label of a specific instrument in the output such as violin or piano.

Domain:

The domain of one variable encompasses the set of all instrument types that could possibly have been used in the dataset such as ({violin, piano, guitar etc.})

Constraints:

- **Feature Matching:** It is required that the features drawn from the instruments fits within ranges that are established for each category type of instruments.
- **Temporal Consistency:** The sequential ordering of patterns as recognized by the LSTM should correspond to the sound structure expected of the instrument in question.
- **Noise Tolerance:** The process of classification should remain unaffected by minor distortions or other background noise found in the audio file.
- **Uniqueness:** This implies one audio file can never fall into more than instrument category.

Solution Strategy 1:

1. **Feature Extraction and Preprocessing :**
   - First step consists of standardization of the audio data, this is done by extracting MFCC features.
   - Ensure that data that will be utilized is clean, does not have any noise as well as inconsistencies.

2. Classification Engine as CSP Solver:

- Use a CNN LSTM based model for the assignment of instrument labels on the audio clips.
- Cnn Layers: The first substep involves defining the spatial constraints within the feature space.
- Enforce temporal constraints through the retention of sound sequentials.
- During both the training and the performing of the inference task, the model works as a constraint solver and professors temporal coherence, feature and other important factors.

3. Optimisation for CSP Solution optimum with respect the imposed constraints and costs is reached. finding a weighted estimation so as to search for a reasonable solution that covers all constraints. Ensures statistical consistency is observed by minimizing the total error during gradient based optimization.

4. **Validation –** the test the proposed solutions by performing classification accuracy on the test dataset and see whether all the specified constraints have been met across all of the samples.

   This CSP framework guarantees that the system is in compliance with the classification rules, while at the same time taking advantage of sophisticated AI applications in order to take account of real-world factors and provide credible outcomes.

Chapter 5

Results

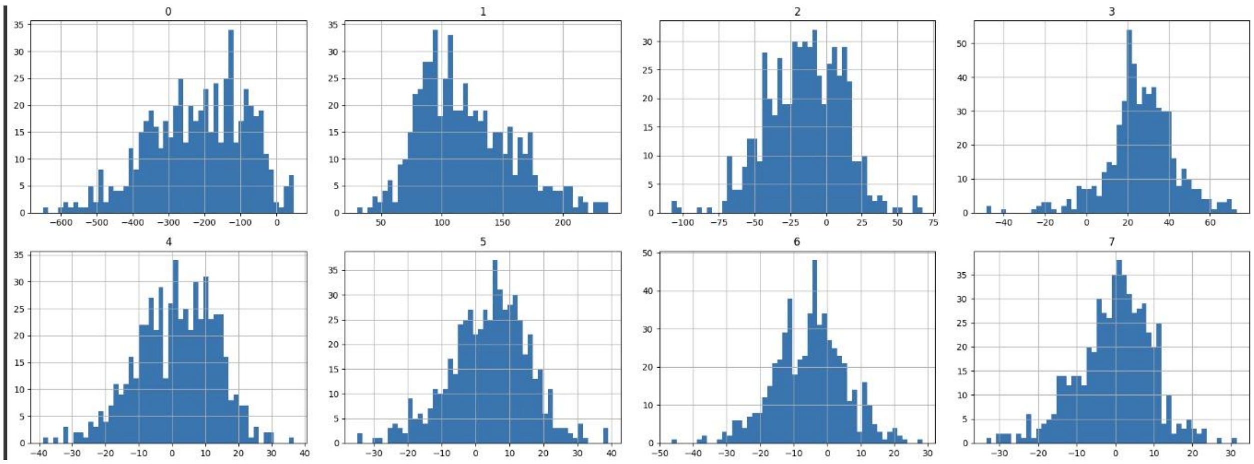

Fig 1: Mel-Frequency Cepstral Coefficient (MFCC) distributions

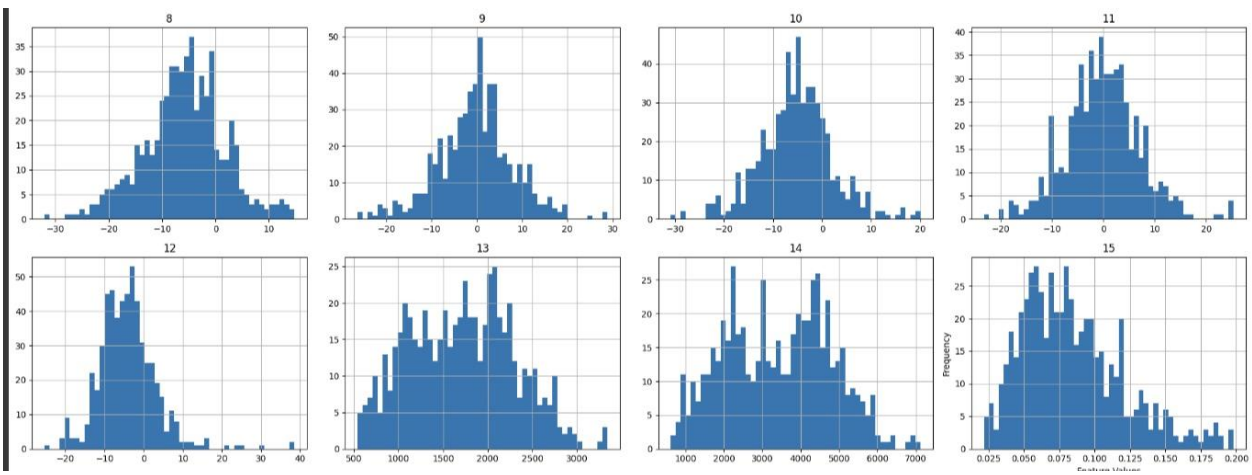

Fig 2: Mel-Frequency Cepstral Coefficient (MFCC) distributions

Images demonstrate visualizations of Mel-Frequency Cepstral Coefficient (MFCC) distributions, which are among the most commonly used features in audio and speech processing. The MFCC coefficients describe the spectral characteristics of audio signals.

Important observations:

- The distributions are different in spikiness, bimodality, and skewness, reflecting the diversity of underlying audio signals.

- The values ranges and amounts of MFCC values change over plots, therefore this may imply that the audio signals may be coming from possibly different sources, environments, or conditions for processing.
- Some of the MFCC coefficients whose distributions appear fairly distinct and separable as such 2, 3, 6, 8-11, so perhaps useful in audio classification.
- Different distributions of the MFCC can be related to factors of type of sound, speaker/instrument characteristics and the recording conditions.

Without additional context, it seems that the MFCC visualizations described above might be helpful for tasks such as audio classification, speech recognition, music analysis, or sound event detection.

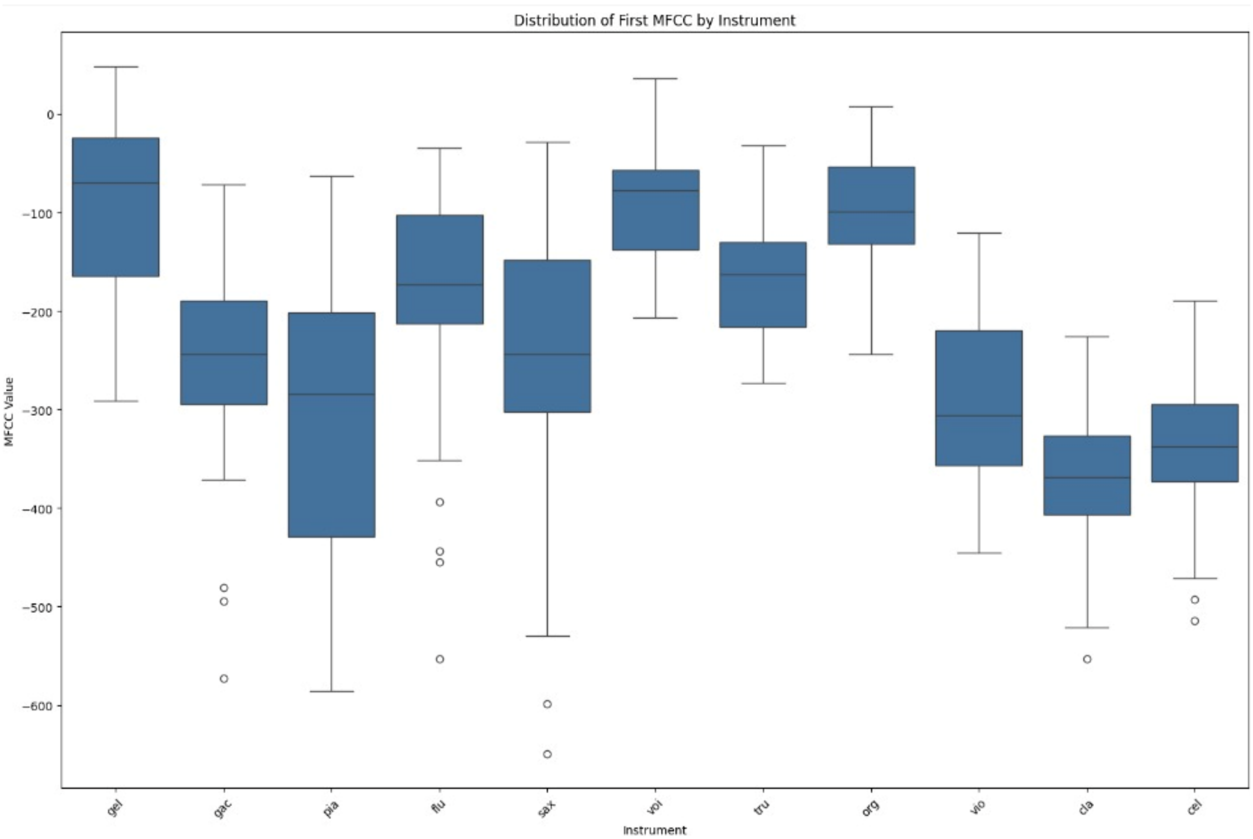

Fig 3: Distribution of the First MFCC by Instrument

KeyObservations:

1. Spread and Central Tendency: The median in each box represents the central tendency of the MFCC values for each instrument. For example, "flu" or flute has a much higher median than "cel" or cello.

2. Outliers: Those small dots below the whiskers indicate that for some of the instruments, some of its sounds are definitely outliers, since they lie below the normal range of MFCC.

3. Application to the Project: From these distributions, we can see how differently or overlapped MFCC characteristics between instruments are and how that affects the accuracy of classification. Tightly clustered distributions indicate better separability for some instruments.

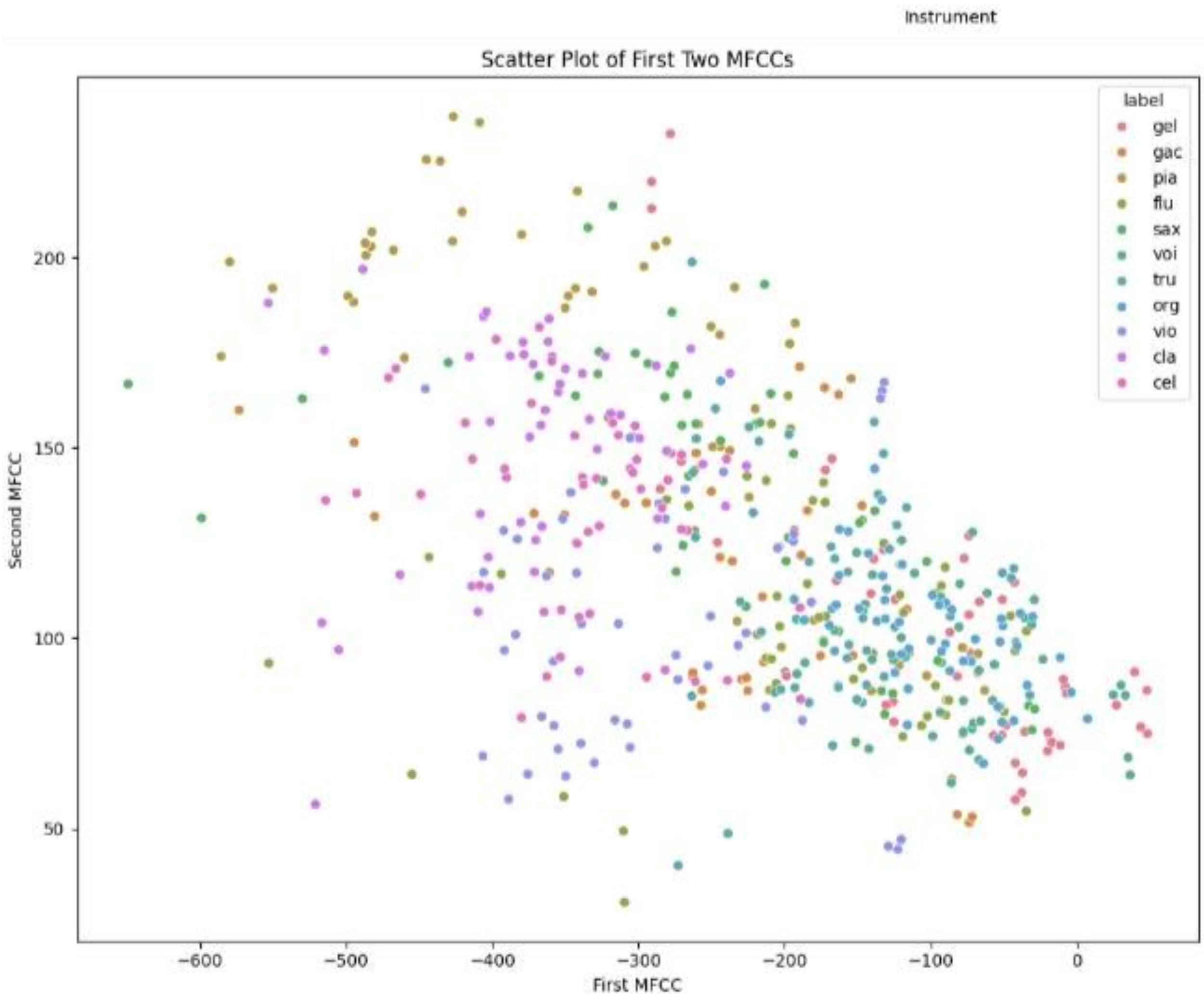

Fig 4: First Two MFCCs

Purpose: This plot depicts a relationship between the first two MFCC coefficients for frames of audio.

Key Observations:

Clusters: Most of the instruments exhibit clustering patterns, like "pia" (piano) or "org" (organ); while others overlap and produce different levels of distinguishability.

Data Distribution: The distribution on both axes presents the range of range of the variations included by these features.

Application to the project: This visualization would be really useful to interpret feature space in dimensionality reduction techniques or feature engineering for ML models.

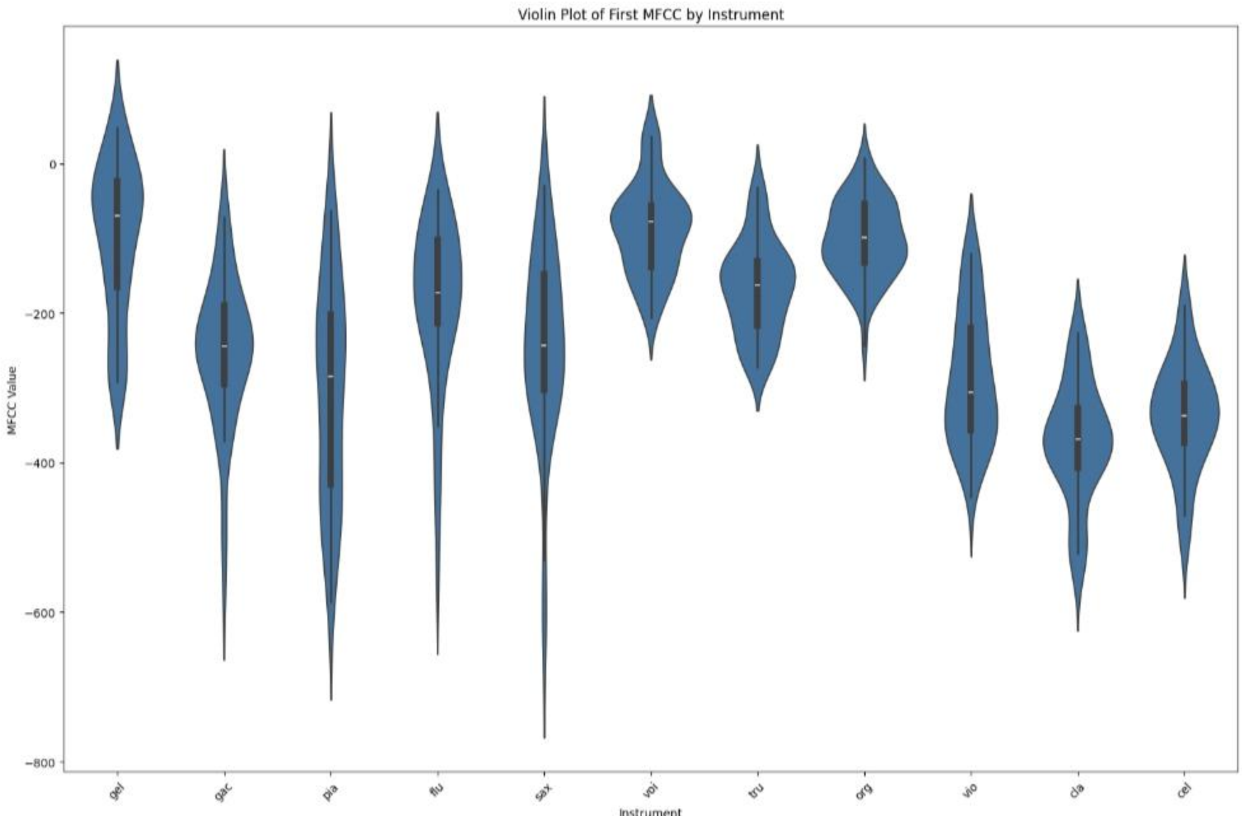

Fig 5: Violin Plot: First MFCC by Instrument

This presents an alternate view of the distribution of first MFCC coefficient for every instrument.

Key Observations:

Density Estimation: The "violin" shape is the probability density function of the MFCC values, which provides a more complete view of the distribution than the box plot.

Central Tendency and Spread: The median and interquartile range are still visible, so it is possible to compare central tendency and variability between instruments.

Application to the Project: This visualization supplements the box plot and provides more insight into the statistical properties of the MFCC features and how they differ between musical instruments.

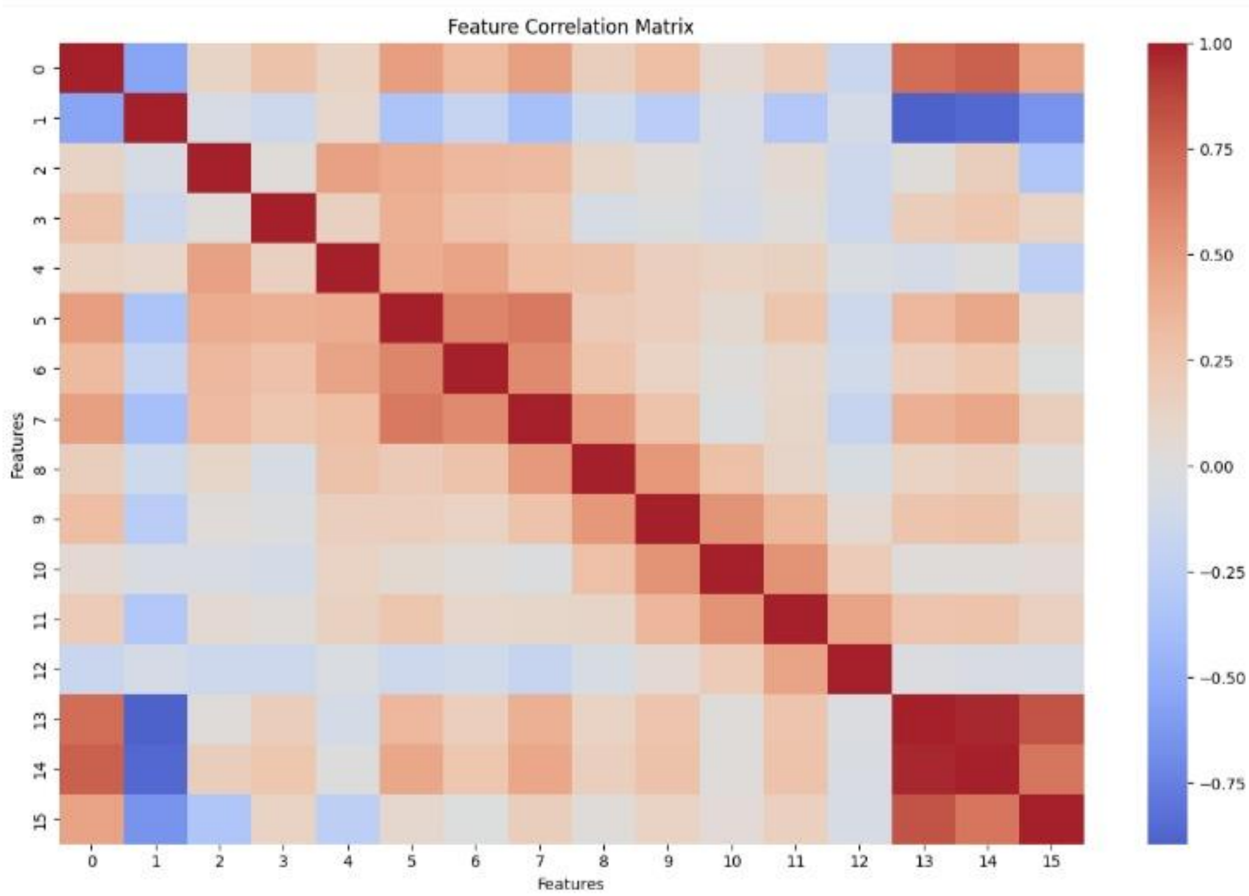

Fig 6: Feature Correlation Matrix

Purpose: Plots all the different MFCC coefficients in pairwise correlation.

Key Findings:

Highly Correlated Features: The cells colored in red near to the diagonal are highly correlated features, which essentially means redundancy in data.

Weak Correlations: Negative or independent features are marked in blue colour.

Application to the Project: These correlations help in feature selection or transformation. Features with low correlation would be favored in reducing dimensionality to ensure diversity in the model's input.

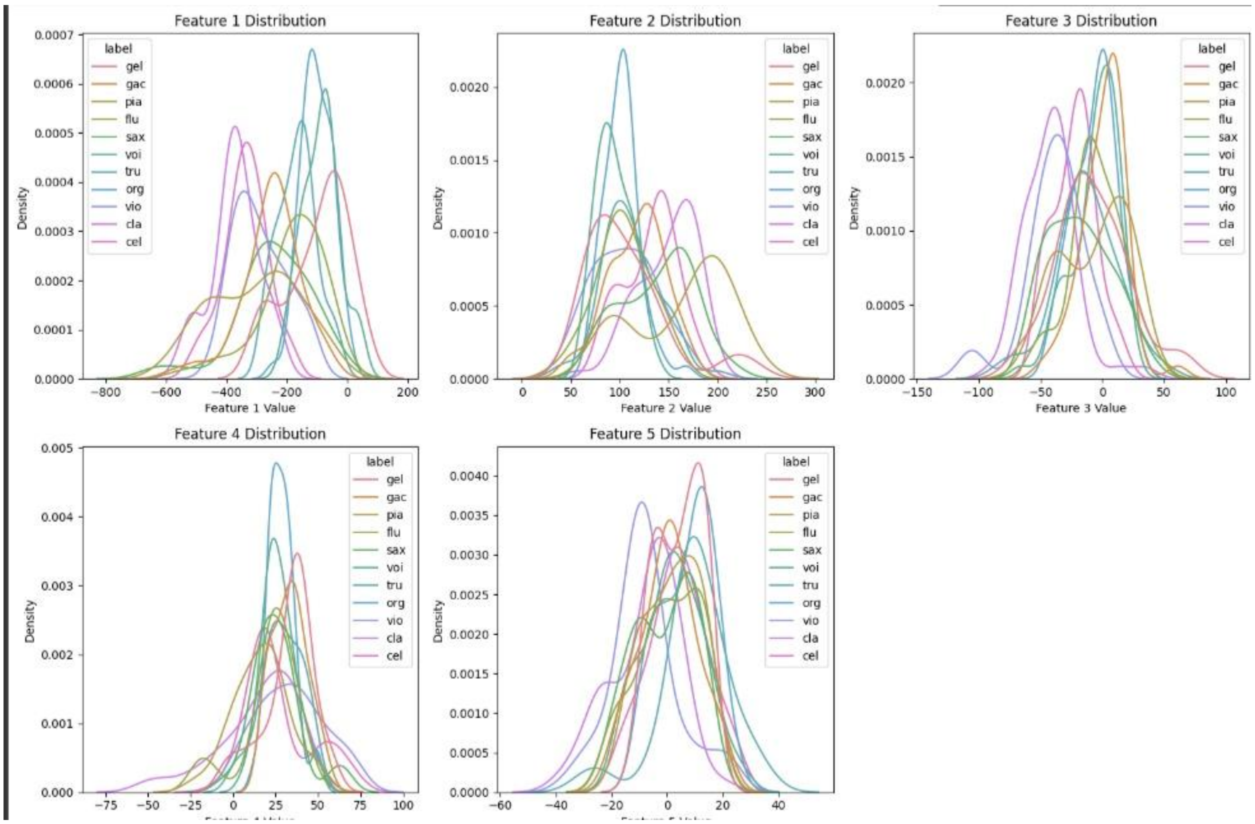

Fig 7: Feature Distributions

Purpose: Display the individual distributions of MFCC features for 1 to 5 from each instrument.

Key Observations:

Distinct Distributions: Shape and spread of the distributions differ across instruments, which means that the spectral characteristics they capture differ.

Overlap and Separability: Some distributions overlap more while others are more distinct. This could indicate a higher separability of certain instruments.

These visualizations can be used for guiding the design of the classification model architecture so as to better capture the unique spectral patterns of each instrument as well as inform feature selection and weighting.

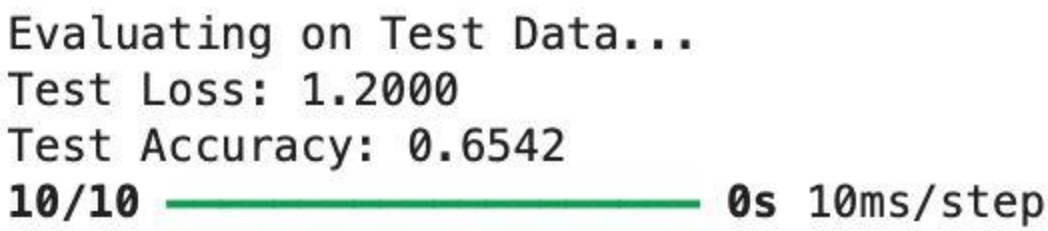

Fig 8: LSTM Results

The image sums up how the model is performing over test data:

- Test Loss: 1.2000
- Test Accuracy: 0.6542
- Evaluation Progress: 10/10 (complete)
- Inference Speed: 0s 10ms/step

Thus, the test loss of the model appears to be 1.2000 and test accuracy 65.42% on the test data set. The evaluation is complete, and the model is able to do inference at the speed of 10 milliseconds per step. These metrics indicate the overall efficiency and effectiveness of the trained machine learning model.

Fig 9: LSTM Results

The model has a training accuracy of 67.34% but a lower test accuracy of 54.64%, indicating potential overfitting. The suboptimal test performance suggests the need for further model refinement to improve generalization.

Chapter 6
Conclusion and Future Scope

The applicability of AI techniques particularly in CNN-LSTM architectures for the classification of the audio data into categories corresponding to different musical instruments is hereby revealed by the Harmony Hub project. Exploiting the MFCC feature extraction system, the system thereby captures the spectral and temporal characteristics of the different instruments and so has a high accuracy level of classification. The hybrid model is very suitable for audio-based applications due to the spatial pattern recognition as well as sequential analysis involved. This project demonstrates the power of AI in modifying how instruments are learned. This is a robust system for use by students and hobbyists to determine their capabilities with real-time feedback.

Future Scope:

This project would allow much more scope for a large set of instruments' creation, with much musical taste and genres. This can give very good classifications even with minimal training datasets available, if properly introduced with advanced techniques in transfer learning or attention. This can get extended to providing comments in playing technique or even quality of performance, making it the ultimate music tutoring application.

This also can be used in interactive sessions through real-time audio input such as playback of live instruments. In this manner, the system can prove to be suitable for learners both at a beginner and a professional level. With deployment on mobile and web, Harmony Hub will reach the masses and hence be available to the larger population thus making it more accessible and inclusive for music education.

References

1. https://link.springer.com/book/10.1007/978-3-319-21945-5
2. https://link.springer.com/chapter/10.1007/978-3-662-55004-5_39
3. **https://zenodo.org/records/1290750#.WzCwSRyxXMU**
4. **https://ieeexplore.ieee.org/document/7952267**