Music Instrument Recognition Using Machine Learning

Presented by Group 1

Diya Dineep - CB.SC.U4AIE24110

Vemuri Monisha Reddy - CB.SC.U4AIE24157

Introduction

- Music Instrument Recognition (MIR) is a machine-learning task that identifies musical instruments from audio recordings.
- This technique is commonly used in applications like music transcription, sound classification, and automatic tagging in music streaming services.
- In this project, we employ Support Vector Machine (SVM), a widely used supervised learning algorithm, to classify different musical instruments based on extracted audio features.
- Support Vector Machines (SVM) are widely used in this task because of their efficiency in handling high-dimensional data and their ability to generalize well with limited training data.

SVM (Support Vector Machine)

What ?

Why Only SVM?

How?

Relevance

- Many studies use deep learning techniques such as CNNs and RNNs, but we focus on SVM which is computationally more efficient.
- Services like Spotify and Apple Music can use instrument recognition to improve playlist recommendations and automatic tagging of songs.
- Unlike deep learning, SVM can run efficiently on low-power hardware (Raspberry Pi, mobile devices), making it useful for real-time applications.

Advantages

- Improved Accuracy with SVM: Works well for high-dimensional feature spaces.
- Simplified Processing with Monophonic Music: Easier classification compared to polyphonic music.
- Scalability: Can be trained with more instrument types for improved recognition.
- Real-time Applications: Useful for live music analysis and recommendation systems.

Objectives

- To develop an SVM-based model for musical instrument recognition in monophonic recordings.
- To train and test the model on monophonic instrument datasets and evaluate performance using accuracy, precision, recall, and F1-score.
- To improve classification accuracy by optimizing SVM kernel functions and hyperparameters.
- To enable applications such as music classification, transcription, and sound retrieval.

Literature Review 1

Exploiting cepstral coefficients and CNN for efficient musical instrument classifcation

Methodology

- A CNN model is built and trained to classify 20 different musical instrument classes.
- Extract Mel-Frequency Cepstral Coefficients (MFCCs) from the instrument sounds. These features help the model recognize the timbral properties of instruments.
- The CNN model is trained on the combined dataset (Philharmonia + UIOWA).
- The trained model achieves state-of-the-art accuracy in instrument classification.

Advantages

- Highly Accurate – Achieves state-of-the-art results using CNN.
- Feature-Based Approach – Uses MFCCs, which are effective for audio classification.
- Useful for AI Applications – Can be used in music transcription, recommendation systems, and audio analysis tools.

Literature Review 1

Metrics

- **Accuracy - Achieved 98% accuracy on the test set.**
- **Mean F1 score - 97.5%**

Limitations

- **Augmentation techniques may not fully represent original audio characteristics.**
- **Limited feature extraction methods.**

Literature Review 2

Music Instrument Recognition using Machine Learning Algorithms

Methodology

- The paper focuses on music instrument detection from an audio clip using machine learning algorithms like ANN and CNN.
- ANN and CNN are used for feature based classification and analysing.
- The dataset includes audio samples of eight instruments.
- The ANN and CNN models classify audio into one of the eight predefined instrument classes with high accuracy.

Advantages

- AI-Based Recognition – Uses advanced neural networks to automate the process.
- Handles complex sounds – Can identify instruments even when they are mixed sounds.
- Fast Processing – Works faster than manual identification by humans.

Literature Review 2

Metrics

- **Accuracy - CNN model shows more accuracy compared to ANN.**
- **Confusion matrix – CNN is relatively better compared to ANN.**

Limitations

- **CNN requires significantly more time for training and testing compared to ANN due to its deeper architecture and computational requirements.**
- **CNNs require large, well-annotate datasets for effective training .**

Literature Review 3

Musical Instrument Identification Using Machine Learning

Methodology

- The paper focuses on audio recordings with one instrument playing one pitch and identifying them using machine learning.
- The study focuses on viola, piano, and ukulele recordings with single-pitch sounds.
- Differences in instrument sounds are analyzed using harmonic frequency content from spectrograms.
- The K-Nearest Neighbors (KNN) algorithm is used for classification.

Advantages

- Simple and Efficient – Uses KNN, which is easy to implement and works well for small datasets.
- Accurate Identification – Achieves 80% accuracy in classifying instruments.
- Useful for Streaming Services – Can assist platforms in providing personalized recommendations.

Literature Review 3

Metrics

- Accuracy - the study achieved 80% accuracy.
- Confusion matrix - Helps identify errors like false positives (wrongly classified as another instrument) and false negatives (missed correct classifications).

Limitations

- Single instrument & single pitch only
- Not suitable for real-world noisy environments

Literature Review 4

Classification of Musical Instruments' Sound using kNN and CNN

Methodology

- **Automate the identification and categorization of instruments based on their acoustic characteristics.**
- **Machine learning algorithms used -**

1. **k-Nearest Neighbors (kNN) – A simple algorithm that classifies instruments based on similarity.**
2. **Convolutional Neural Network (CNN) – A deep learning model that recognizes spectrogram patterns.**

- **The models are trained and tested to check their performance.**

Advantages

- **Highly Accurate Models – kNN achieves 96% accuracy, and CNN achieves 88% accuracy.**
- **Multiple ML Models Used – Uses both traditional (kNN) and deep learning (CNN) approaches.**
- **Improves Music Analysis – Can be used for instrument recognition in music production.**

Literature Review 4

Metrics

1. **Accuracy -**
- **kNN Accuracy: 96%**
- **CNN Accuracy: 88%**
2. **F1-Score**
- **kNN F1-Score: 0.96**
- **CNN F1-Score: 0.93**

Limitations

- **The model has a relatively small dataset which may limit the model's ability to generalize to other instruments.**
- **The performance of the model depends heavily on the quality of the audio samples.**
- **the current feature extraction method may not capture all relevant acoustic features.**

Research Gap

- Limited feature extraction methods
- Limited dataset size and diversity
- Dependence on high-quality audio samples

Problem Statements

- Why is Instrument Recognition Challenging?
  - Variations in timbre, pitch, and dynamics across different recordings.
  - Difficulty in differentiating similar instruments (e.g., violin vs. viola).
  - Noise and recording quality affect feature extraction accuracy.
- Why Use Monophonic Music?
  - Easier to classify compared to polyphonic recordings, where multiple instruments play simultaneously.
  - Reduces computational complexity and improves recognition accuracy.
  - Useful for automatic transcription and real-time music applications.

Methodology

Step 1: Data Collection (Dataset Selection)

- Choose a dataset containing monophonic music samples (single instrument per audio clip).

Step 2: Preprocessing the Audio

- Convert MP3 to WAV format.
- Convert audio to mono (if stereo).
- Normalize volume and remove silence.
- Convert audio signals into a uniform sample rate .

Step 3: Feature Extraction

- Extract meaningful features from audio signals that help differentiate instruments.
- Common features used for SVM classification:
- MFCCs (Mel Frequency Cepstral Coefficients)
- Spectrograms
- Chroma Features
- Zero-Crossing Rate (ZCR)
- Spectral centroid

Methodology

Step 4: Train the SVM Model

- **Prepare the dataset by splitting it into training and testing sets.**
- **Use SVM with different kernels (Linear, RBF, Polynomial) to find the best performance.**

Step 5: Prediction

- **The trained SVM model predicts the instrument type from new audio samples.**

Novelty

- Testing different SVM kernels to find the best one for instrument classification.
- Combining unique feature sets (MFCC + Spectral Features) instead of just using standard features.

Dataset

Philharmonia dataseet

- **Founded in 1945, the Philharmonia is a world-class symphony orchestra for the 21st century**
- **It includes all standard orchestral instruments, as well as guitar, mandolin, banjo, and a vast array of different percussion instruments.**
- **The samples are suitable for creating any kind of music, no matter the style.**

IRMAS dataset

- **This dataset includes musical audio excerpts with annotations of the predominant instrument(s) present.**
- **he instruments considered are: cello, clarinet, flute, acoustic guitar, electric guitar, organ, piano, saxophone, trumpet, violin, and human singing voice.**

Proposed Timeline

Week 1:Research & Dataset Collection

- Learn about SVM and audio features (MFCC, Chroma, etc.).
- Collect datasets.

Week 2-3:Feature Etraction

- Use Python to extract features
- Save extracted features as a CSV file for training SVM.
- Conduct an initial data visualization to understand feature differences.

Week 4-5:Implementing SVM and training the model

- Implement Support Vector Machine (SVM) using Scikit-learn.
- Test different SVM kernels (Linear, RBF, Polynomial, Sigmoid).
- Train the model on the dataset and optimize hyperparameters (
- Check for overfitting and adjust model settings if needed.

Week 6-7:Model evaluation and performance analysis

- Test the trained models on unseen audio samples.
- Compare SVM kernel performances using accuracy, precision, recall, and F1-score.
- Identify the best-performing kernel.

Week 8-9:Finalising results and submission

- Improve the model by testing different feature combinations.
- Summarize findings in a detailed report.

Thank you!