Project Title:

CNN-Based Music Instrument Recognition System

Project Mentor – Narendra kumar

Group 2

Introduction: A Basic Idea

Music instrument recognition is the process of automatically identifying musical instruments from audio signals. It is widely used in music streaming platforms, recommendation systems, and audio analysis applications. Traditional methods require manual feature extraction, which is inefficient. Deep learning models like CNN and ResNet18 automatically learn features and improve classification accuracy.

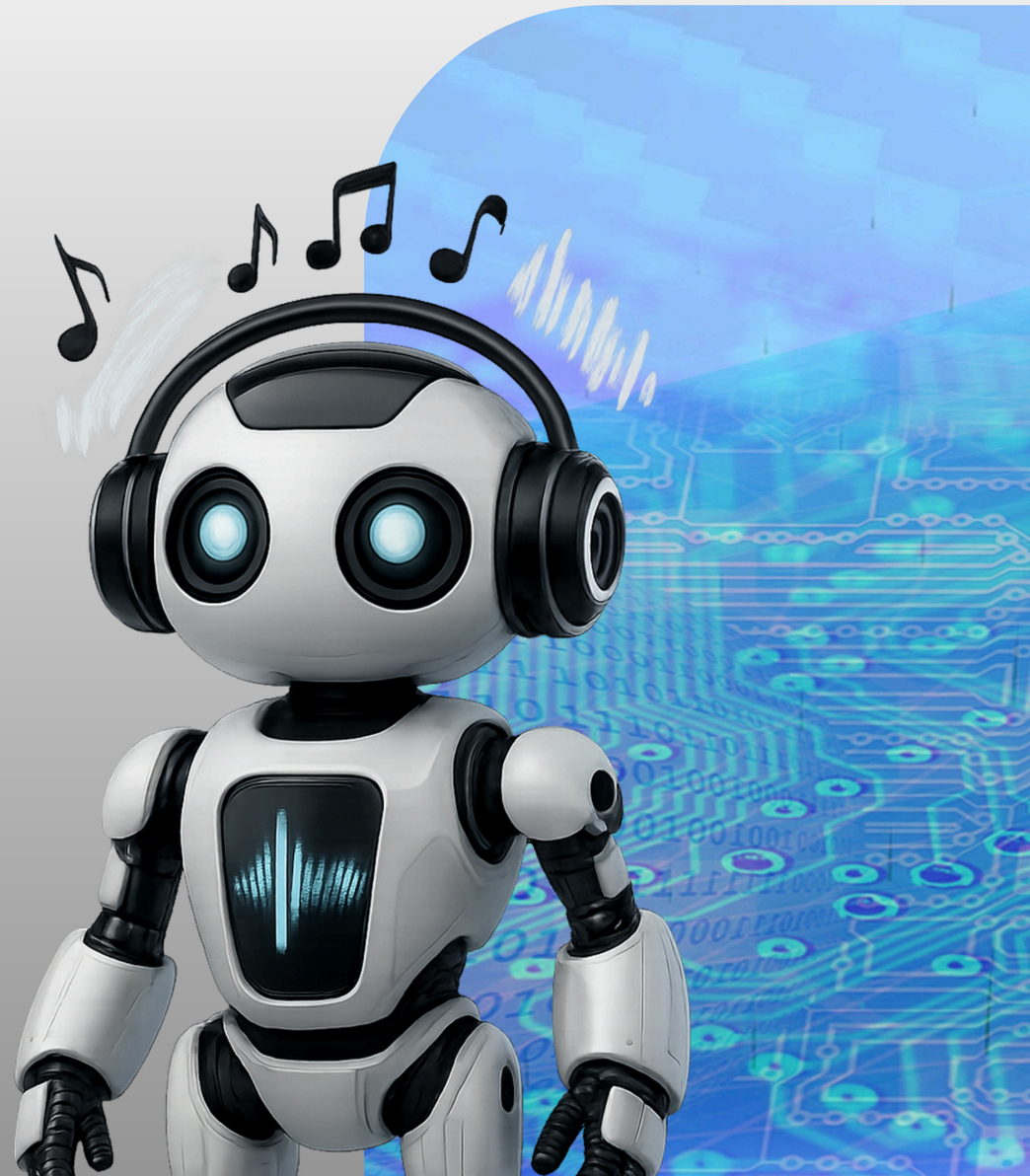
Key Points:

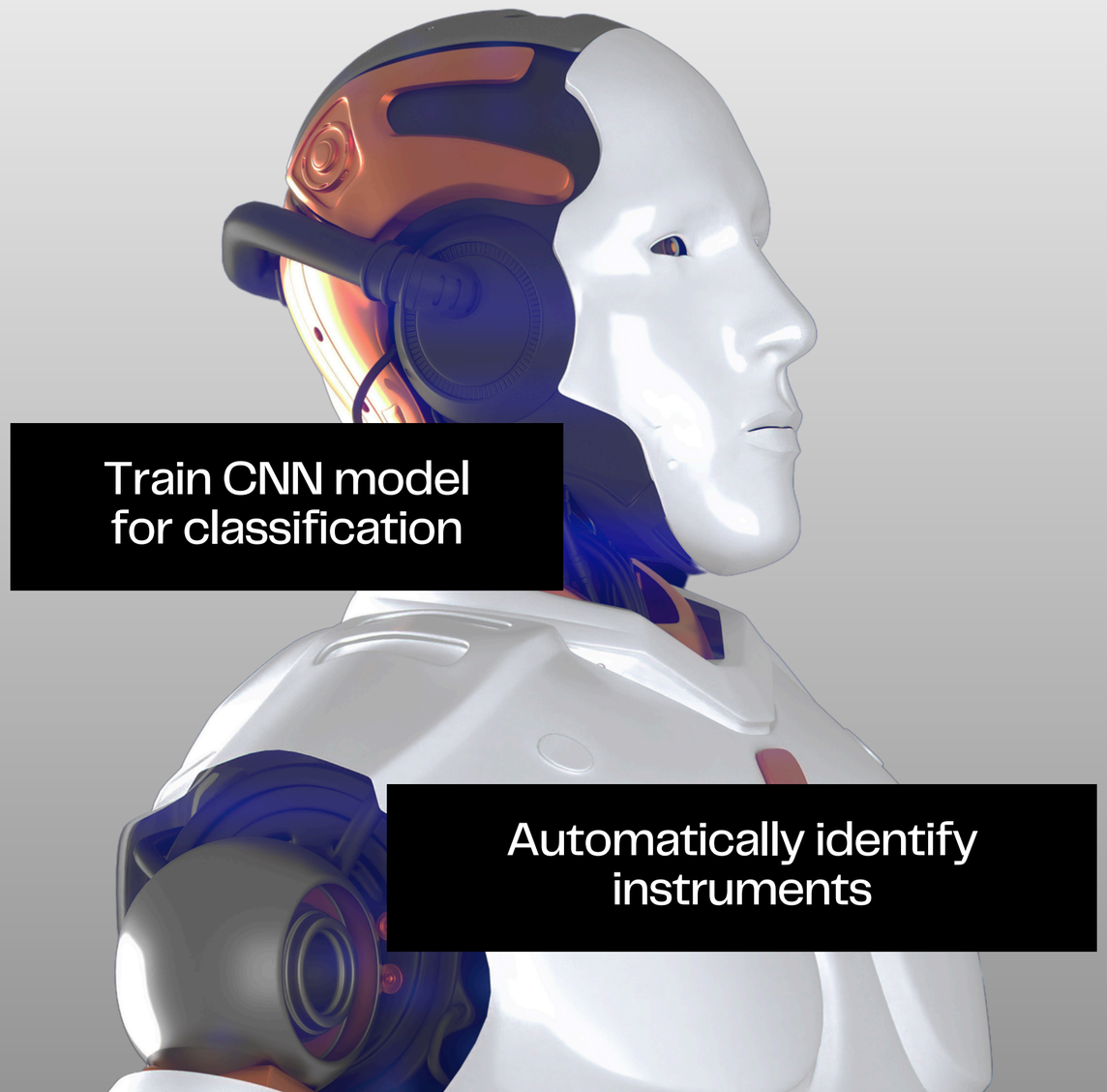
- Automatic instrument identification
- Used in music recommendation systems
- Manual methods are inefficient
- Deep learning improves accuracy

Problem Statement

Identifying musical instruments from raw audio signals is a challenging task because audio signals contain complex frequency patterns, overlapping sounds, and noise. Traditional methods rely on manual feature extraction and tagging, which are inefficient and not scalable. Raw waveform signals do not provide clear patterns for classification, making accurate instrument recognition difficult.

- Audio signals contain overlapping frequencies
- Manual identification is slow and inefficient
- Difficult to extract meaningful features
- Low accuracy with traditional methods





Achieve high prediction accuracy

Train CNN model for classification

Automatically identify instruments

Objective

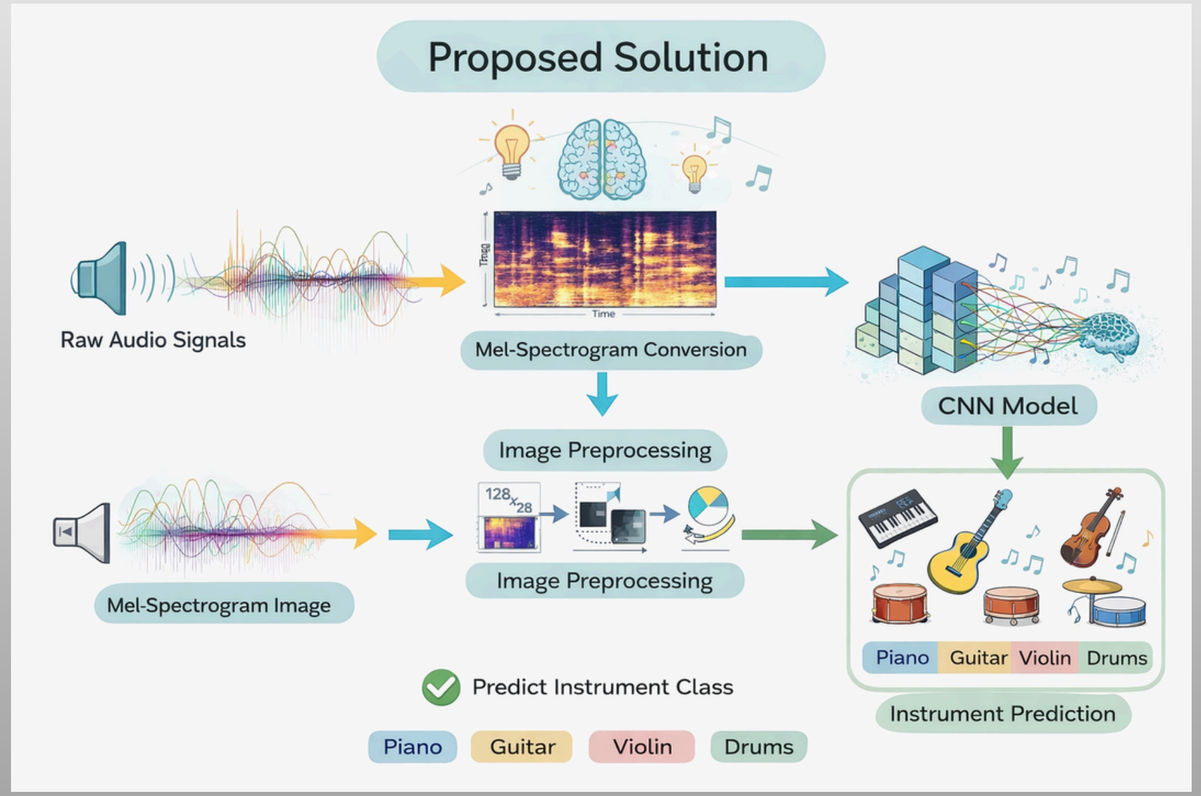
The main objective of this project is to develop a deep learning–based system using Convolutional Neural Networks (CNN) to automatically recognize musical instruments from audio signals. The system converts audio into spectrogram images and trains a CNN model to accurately classify instruments such as piano, guitar, violin, and drums.

Proposed Solution

The system converts audio signals into Mel-Spectrogram images. These images are preprocessed and used as input to a pretrained ResNet18 model. The model extracts features and predicts the musical instrument class.

Steps:

- Audio input
- Spectrogram generation
- Image preprocessing
- ResNet18 feature extraction
- Instrument classification

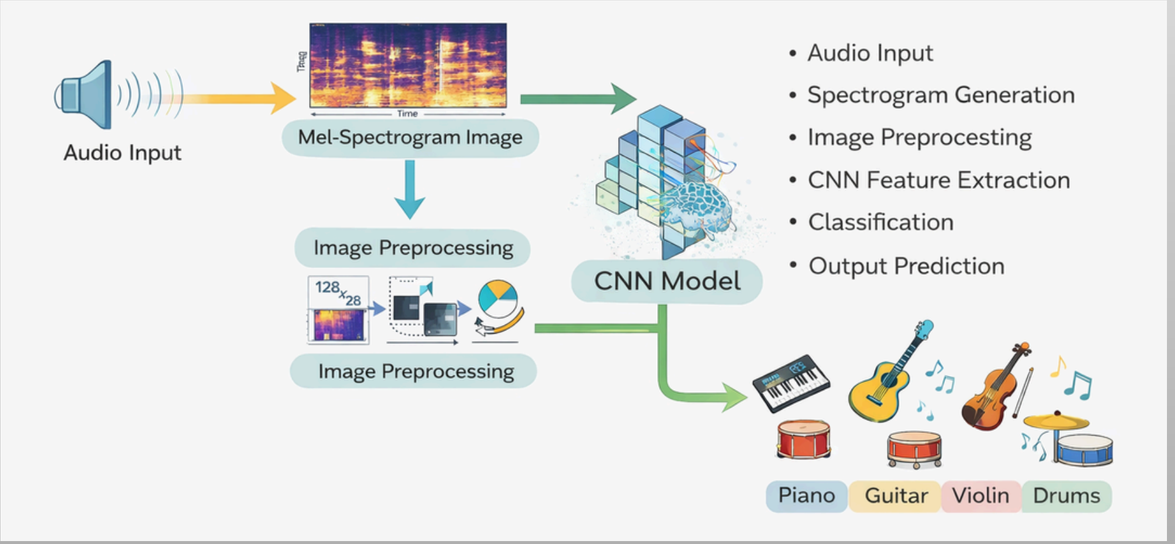


System Architecture

The system follows a pipeline from audio input to instrument prediction. Audio is converted into spectrogram images, preprocessed, and passed into the ResNet18 model for classification.

Steps:

- Audio Input
- Mel-Spectrogram Generation
- Image Preprocessing
- Feature Extraction using ResNet18
- Classification
- Output Prediction

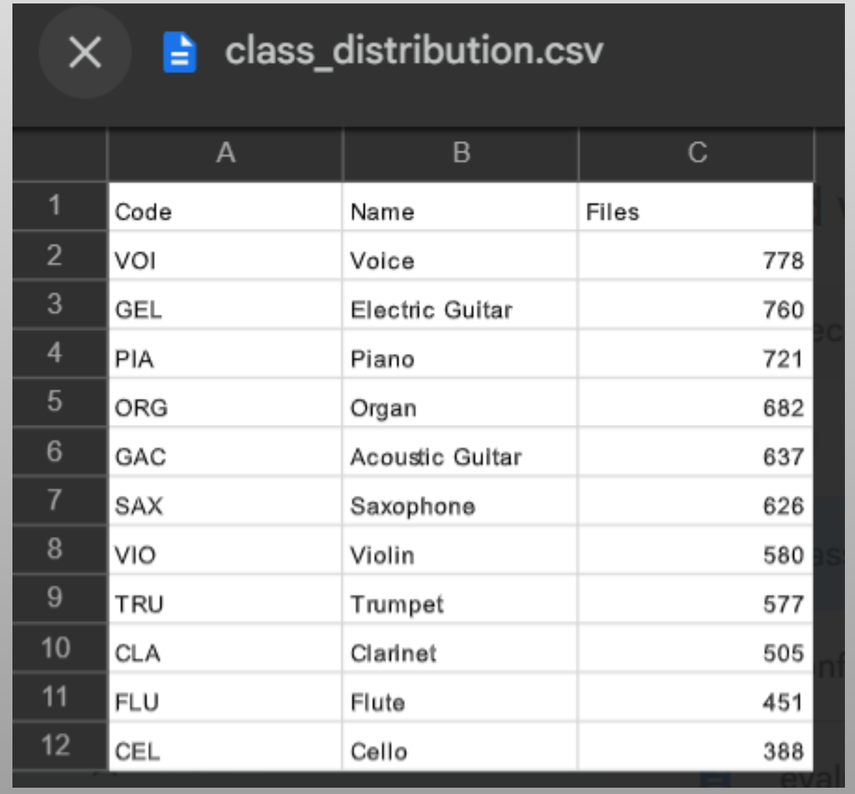


Dataset Overview

IRMAS Dataset

This project uses the IRMAS (Instrument Recognition in Musical Audio Signals) dataset, which contains audio recordings of various musical instruments. The dataset is divided into training, validation, and test sets. Each audio file is converted into spectrogram images and used to train the deep learning model.

- Dataset: IRMAS
- Audio format: WAV
- Instruments: Piano, Guitar, Violin, Drums, etc.
- Training split: 70%
- Validation split: 10%
- Test split: 20%



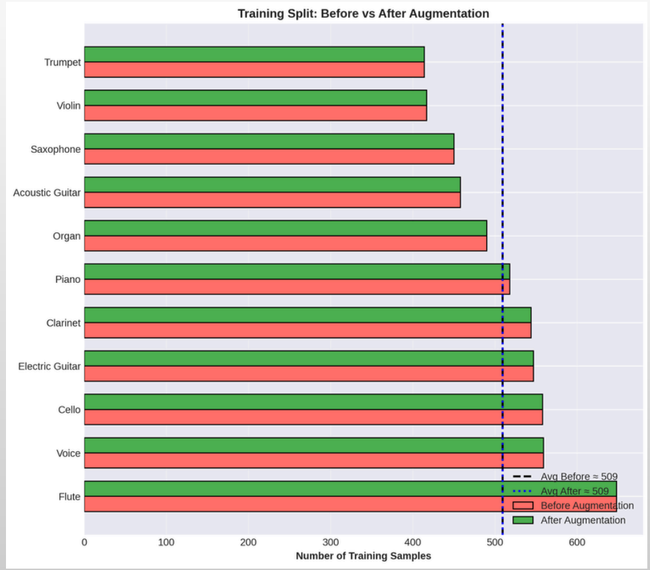
class_distribution.csv

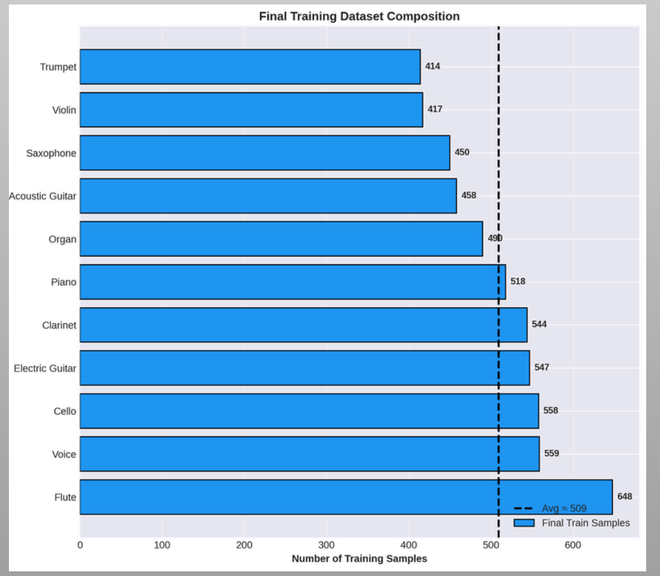
	A	B	C
1	Code	Name	Files
2	VOI	Voice	778
3	GEL	Electric Guitar	760
4	PIA	Piano	721
5	ORG	Organ	682
6	GAC	Acoustic Guitar	637
7	SAX	Saxophone	626
8	VIO	Violin	580
9	TRU	Trumpet	577
10	CLA	Clarinet	505
11	FLU	Flute	451
12	CEL	Cello	388

Data Augmentation

Data augmentation was applied to increase dataset size and improve model generalization. Augmentation helps prevent overfitting and improves classification accuracy.

- Gaussian Noise added
- Time Stretch applied
- Pitch Shift applied
- Audio Shift applied
- Improves model performance





Mel-Spectrogram Generation

Audio files are converted into Mel-Spectrogram images using Librosa library. Spectrogram represents frequency vs time information, which is used as input to the deep learning model.

Configuration:

- Sample rate: 22050 Hz
- Duration: 3 seconds
- Mel bands: 128
- FFT size: 2048

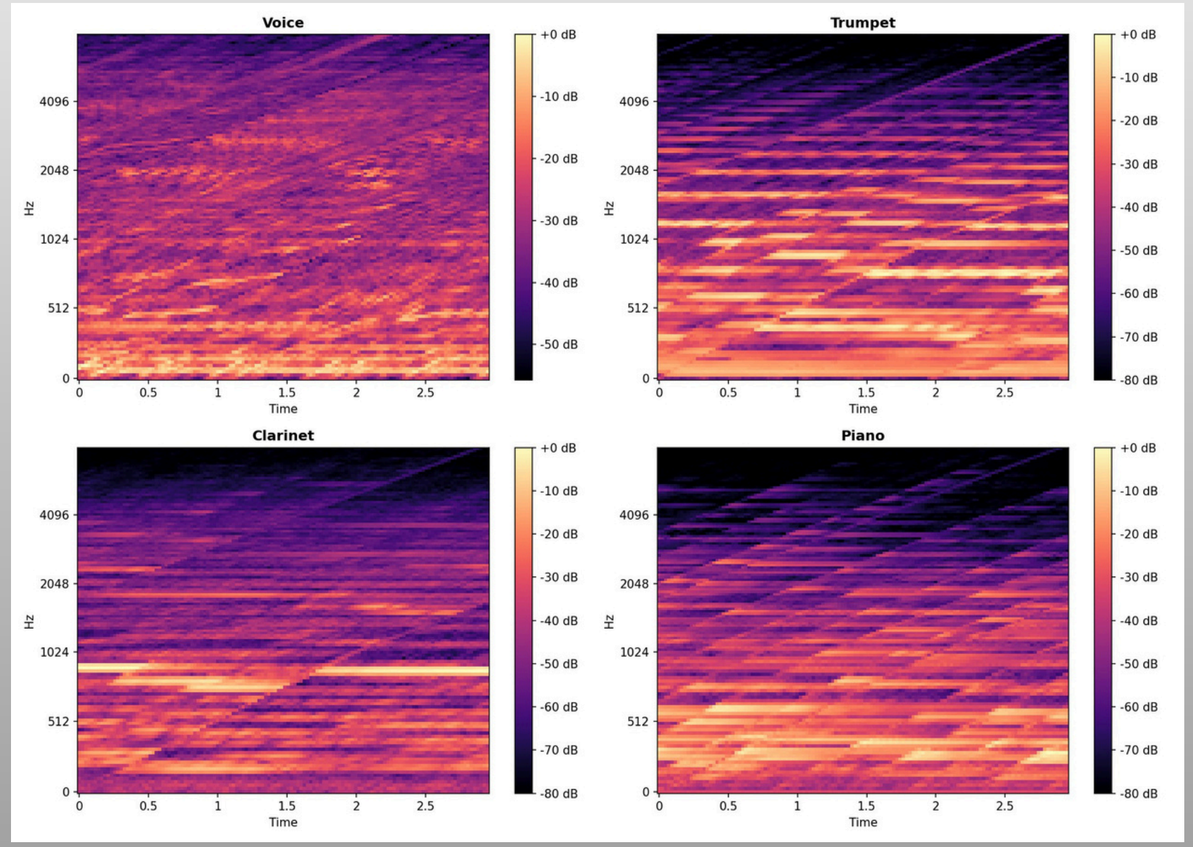
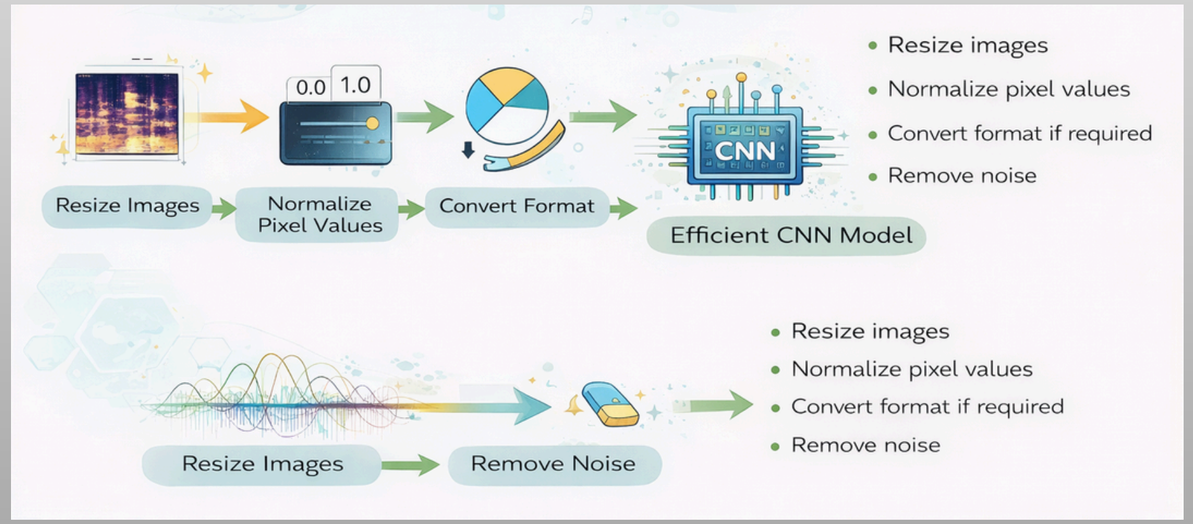


Image Preprocessing

Before training, spectrogram images are preprocessed to improve model performance. Preprocessing ensures uniform size, normalization, and noise reduction, making the CNN model more efficient.

Steps:

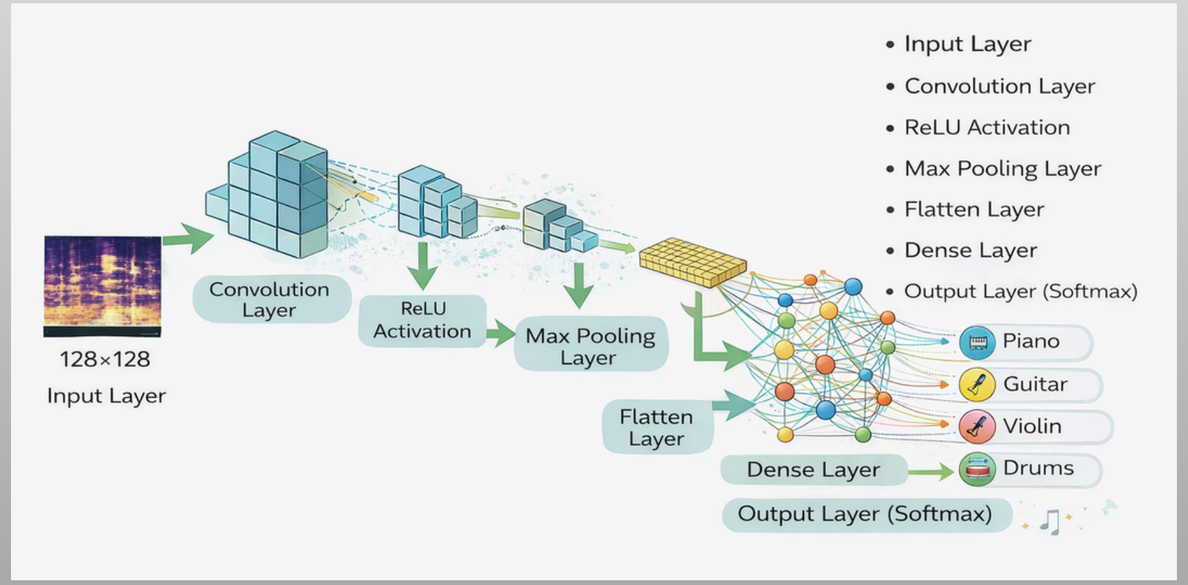
- Resize images
- Normalize pixel values
- Convert format if required
- Remove noise



ResNet18 Architecture

This project uses ResNet18, a deep convolutional neural network pretrained on ImageNet dataset. Transfer learning allows the model to reuse learned features and improves performance. The final layer is modified to classify musical instruments.

- Pretrained ResNet18 model
- Transfer learning used
- Final layer modified for classification
- Uses convolution layers and residual connections
- Improves accuracy and training speed



Training Configuration

The CNN model based on ResNet18 was trained using spectrogram images generated from audio signals. Proper training parameters were selected to ensure stable learning and improve classification accuracy. The model was trained using PyTorch framework with optimized settings.

- Model used: CNN (ResNet18)
- Number of Epochs: 40
- Batch Size: 32
- Optimizer: AdamW
- Loss Function: CrossEntropyLoss
- Framework: PyTorch
- Training performed on CPU/GPU

Training Results

During training, the CNN model learned important features from spectrogram images. The training accuracy improved gradually, and the loss decreased over time. This indicates that the model successfully learned patterns from the dataset.

- Accuracy increased during training
- Loss decreased with each epoch
- Model learned useful features
- Training process was stable
- Good convergence achieved

Evaluation Metrics

The performance of the CNN model was evaluated using standard evaluation metrics. These metrics help measure how accurately the model predicts musical instruments.

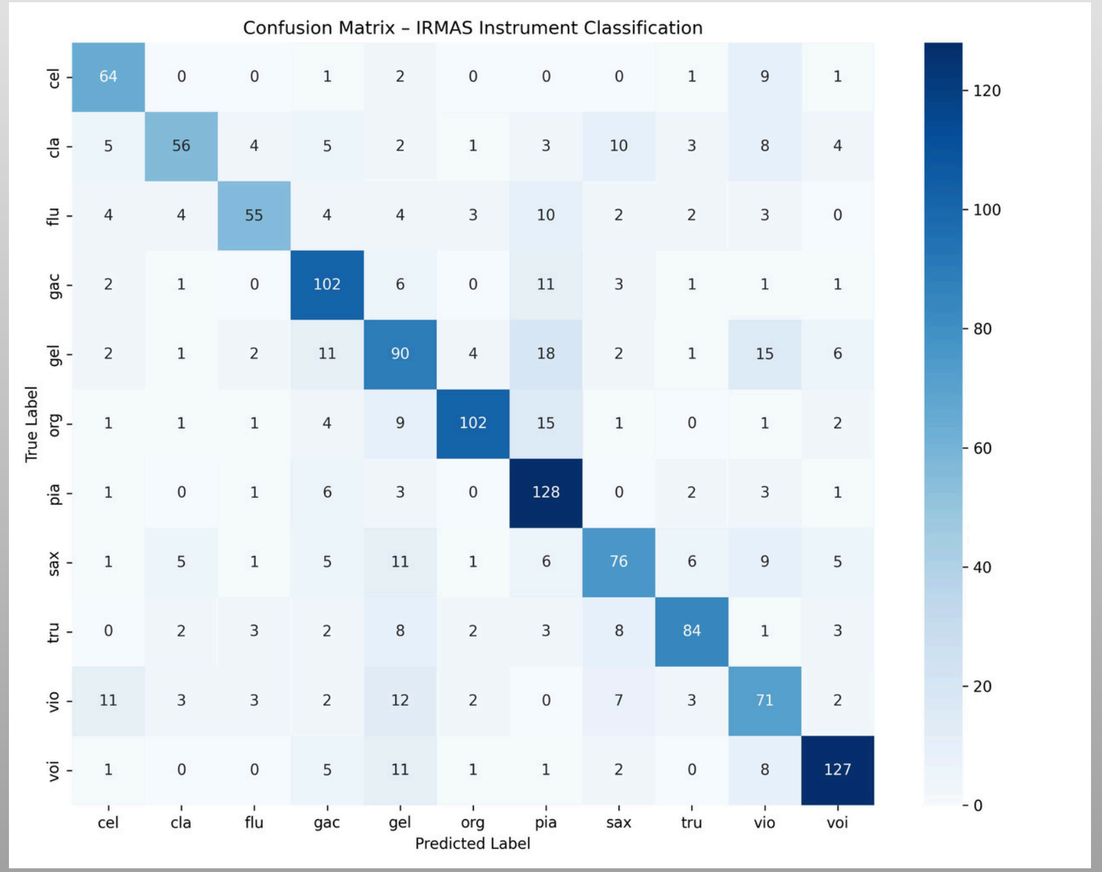
- Accuracy: 0.7095 (70.95%)
- Precision: 0.7202 (72.02%)
- Recall: 0.7095 (70.95%)
- F1 Score: 0.7091 (70.91%)

- Accuracy – measures overall correct predictions
- Precision – measures correct positive predictions
- Recall – measures detection of actual positives
- F1 Score – balanced measure of precision and recall
- Used to evaluate model performance

Confusion Matrix Analysis

The confusion matrix shows the performance of the CNN model for each instrument class. It helps identify correctly classified instruments and misclassifications.

- Shows correct and incorrect predictions
- Helps analyze model performance
- Identifies misclassified instruments
- Useful for improving model accuracy
- Provides detailed classification results



Challenges

Several challenges were faced during the development and training of the CNN model due to dataset limitations and similarities between instrument sounds.

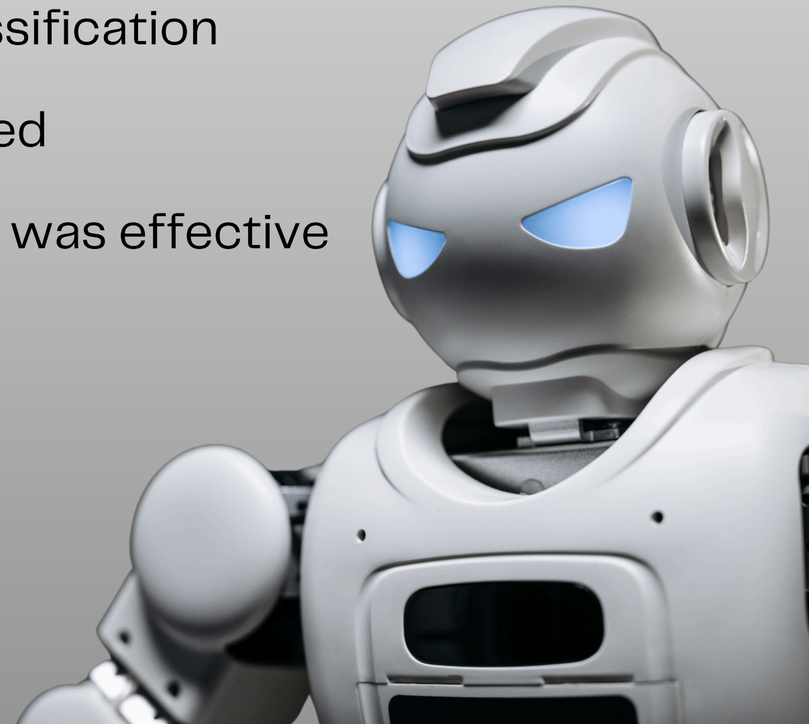
- Class imbalance in dataset
- Similar frequency patterns between instruments
- Noise in audio signals
- Difficulty in feature extraction

Conclusion

The CNN-based ResNet18 model successfully classified musical instruments using spectrogram images. The use of transfer learning and data augmentation improved model performance and accuracy.

- CNN successfully extracted features
- Accurate instrument classification achieved
- Transfer learning improved performance
- Spectrogram conversion was effective

Challenges and Conclusion



The performance of the system can be further improved by using larger datasets and more advanced deep learning architectures.

Future Work

- Increase dataset size
- Improve model accuracy
- Use advanced CNN architectures
- Implement real-time instrument detection
- Improve generalization

THANK YOU