GROUP 11

Music Genre Classification using Deep Learning

- Music Genre Detection
- Convolutional Neural Network
- Recurrent Neural Network
- Conv. Recurrent Neural Networks
- Pre Processing
- Model Training
- Model Validation
- Improvements over Previous Model

OUR TEAM :

MEET OUR MEMBERS

HRISHUBH BHANDARI

DEEP CHAUDHARY

PRAFULLA PATIL

Music Genre Detection

WHAT IS IT?

Music Genre Detection is the process of automatically identifying and categorizing a piece of music into its respective genre(s).

APPLICATION:

- Music Genre Detection is essential for organizing and categorizing large music collections, making it easier to search and navigate through music libraries.
- It can be used in music production and composition, providing insights into the characteristics and elements of different genres.

Convolutional Neural Network

DEEP LEARNING

Convolutional Neural Network (CNN) is a type of deep learning algorithm that is inspired by the human visual system and is well suited for processing multidimensional data, such as images and audio signals.

CNNs have been used to extract meaningful features from audio signals, such as Mel Spectrograms, which are a visual representation of the spectrum of frequencies of a sound over time.

One of the key advantages of CNNs over traditional neural networks is their ability to handle input data of varying sizes. This is because the convolutional layer applies the same filter to the whole input, allowing the network to identify features regardless of the size of the input images or audio signals.

Recurrent Neural Network

DEEP LEARNING

A Recurrent Neural Network (RNN) is a type of neural network that is designed to handle sequential data, such as audio signals. It works by maintaining an internal state, also known as a "hidden state,". The hidden state captures information about the current as well as the inputs that came before it.

One of the main advantages of RNNs is that they can handle variable-length sequences, which makes them well-suited for audio processing tasks.

However, RNNs have some limitations, such as the difficulty in capturing long-term dependencies. These limitations can be addressed by using more advanced RNN architectures, such as Long Short-Term Memory (LSTM) which can help to capture long-term dependencies.

Recurrent Neural Network

DEEP LEARNING

Long Short-Term Memory (LSTM) have additional mechanisms, such as "memory cells" and "gates," that enable them to selectively remember important information and forget irrelevant information.

Additionally, LSTMs have a "memory cell" that can store information over time and selectively update it. This enables the network to selectively remember important information and forget irrelevant information. The ability of LSTMs to capture long-term dependencies in sequential data makes them well-suited for tasks such as speech recognition, natural language processing, and time series prediction.

Convolutional Recurrent Neural Networks

DEEP LEARNING

A Convolutional Recurrent Neural Network (CRNN) is a type of neural network that consists of multiple convolutional layers followed by multiple recurrent layers. The convolutional layers extract features from the audio signal, while the recurrent layers capture temporal dependencies between these features. The output of the recurrent layers is fed into a fully connected layer for final classification.

CRNNs have been successfully used in various audio processing tasks, including speech recognition and music genre classification. They have also been used in other domains, such as image classification and natural language processing.

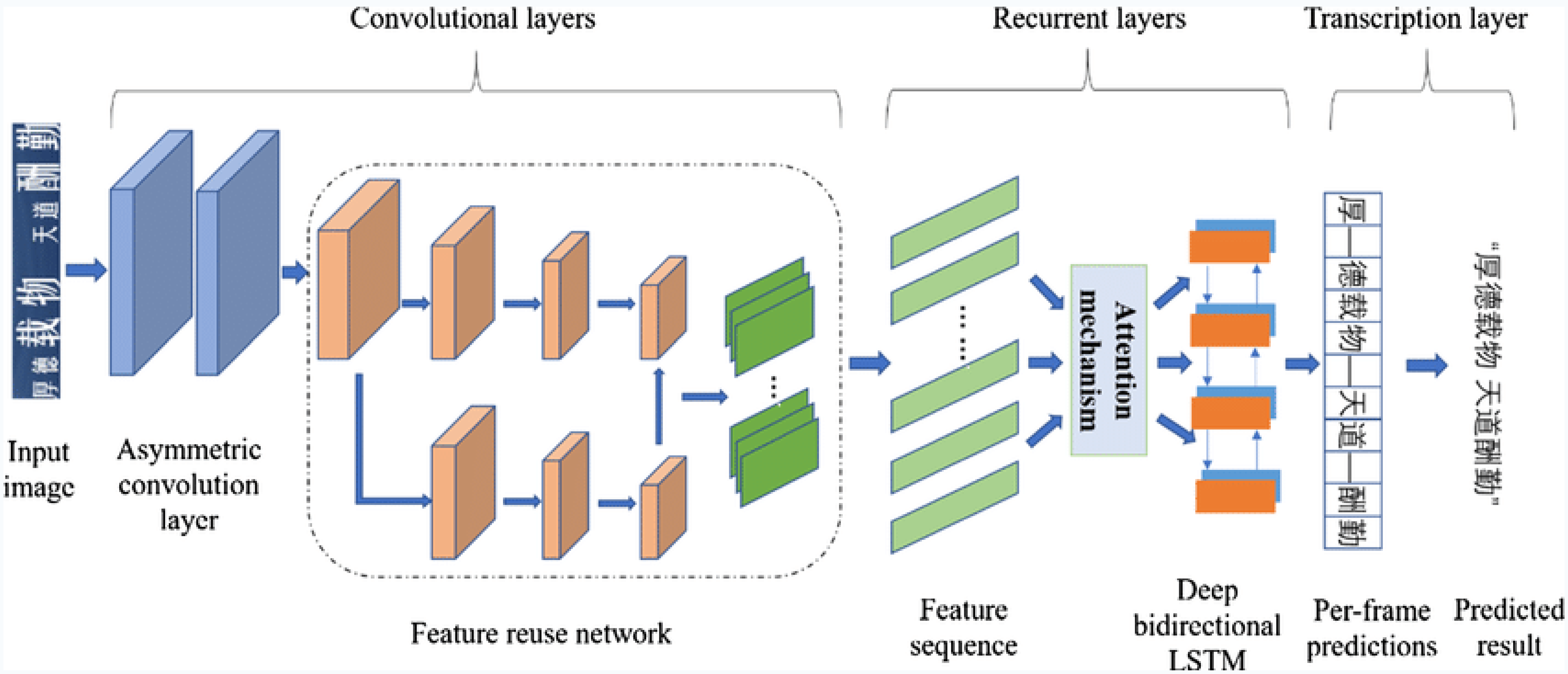


Fig: Example of a Convolutional Recurrent Neural Networks

OUR PROJECT

Implementing a Music Genre Classification using CRNN

GROUP 11

IMPLEMENTATION

1 PRE-PROCESSING

Extracting data from the GTZAN dataset and converting it into Mel Spectrogram to be fed into the CRNN. Similar operations need to be performed on the input as well.

2 TRAINING

Training the Music Genre Classification model to predict the genre of a song based on the Mel Spectrogram of that song.

3 VALIDATION

It involves the validation of the effectiveness of the Music Genre classification on a labeled testset. It gives us a fair idea of the usefulness of the model.

4 PREDICTION

It involves feeding the trained CRNN with an unlabeled input song to be labeled on the basis of the trained model.

STAGE 1:

PRE PROCESSING

MULTIFRAMING:

Multiframe processing for Mel Spectrograms is the technique of dividing an audio signal into multiple frames, computing the mel spectrogram for each frame, and then concatenating the spectrograms to create a multi-frame representation of the audio signal.

CREATING A MEL-SPECTROGRAM

It is the process of conversion of multiple frames of a song into their corresponding Mel Spectrograms to feed the Deep Learning Model. This involves dividing the signal into small overlapping windows and computing the Fourier transform for each window.

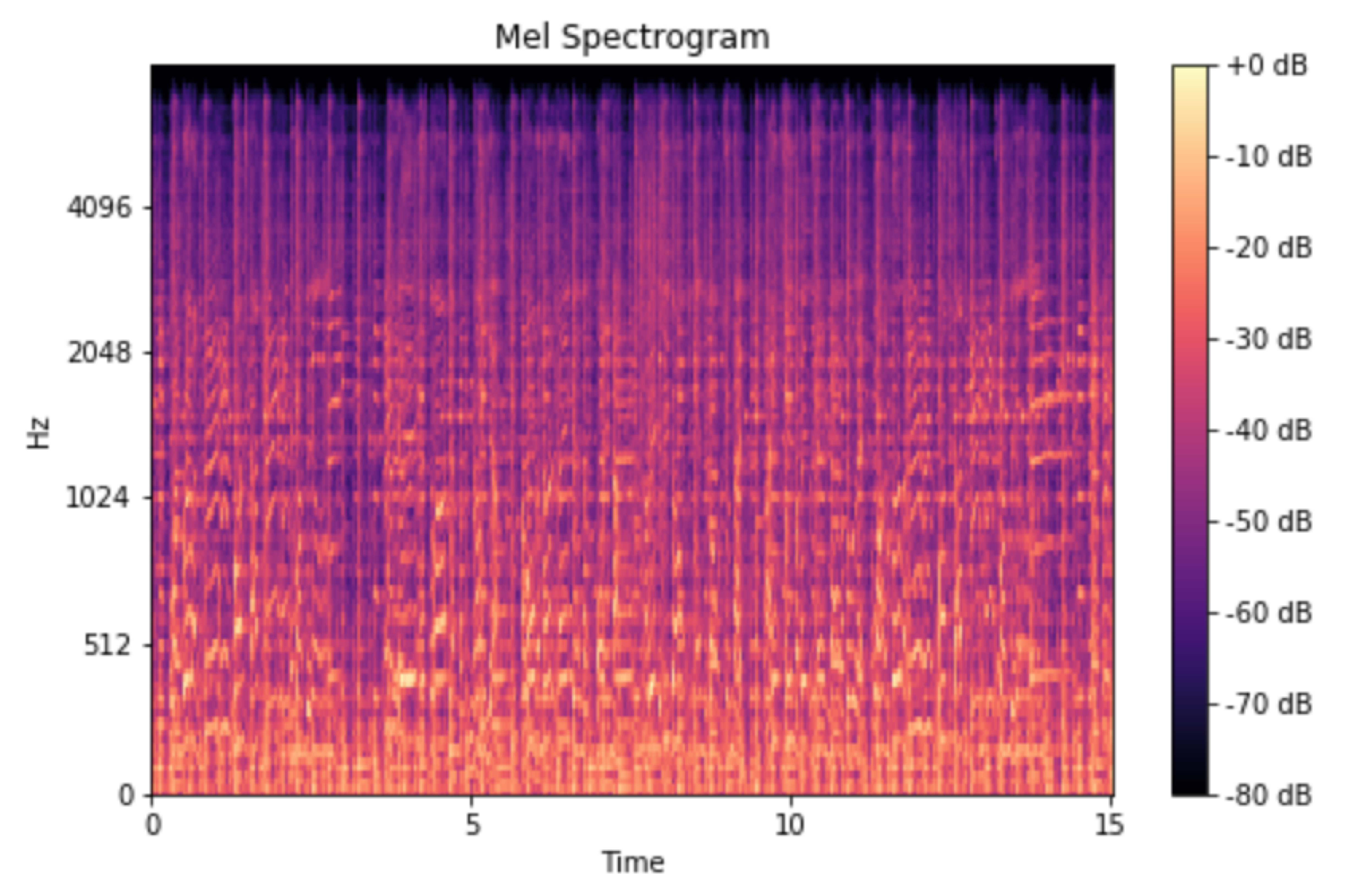


Fig: A Simple Mel-Spectrogram

STAGE 2:

MODEL TRAINING

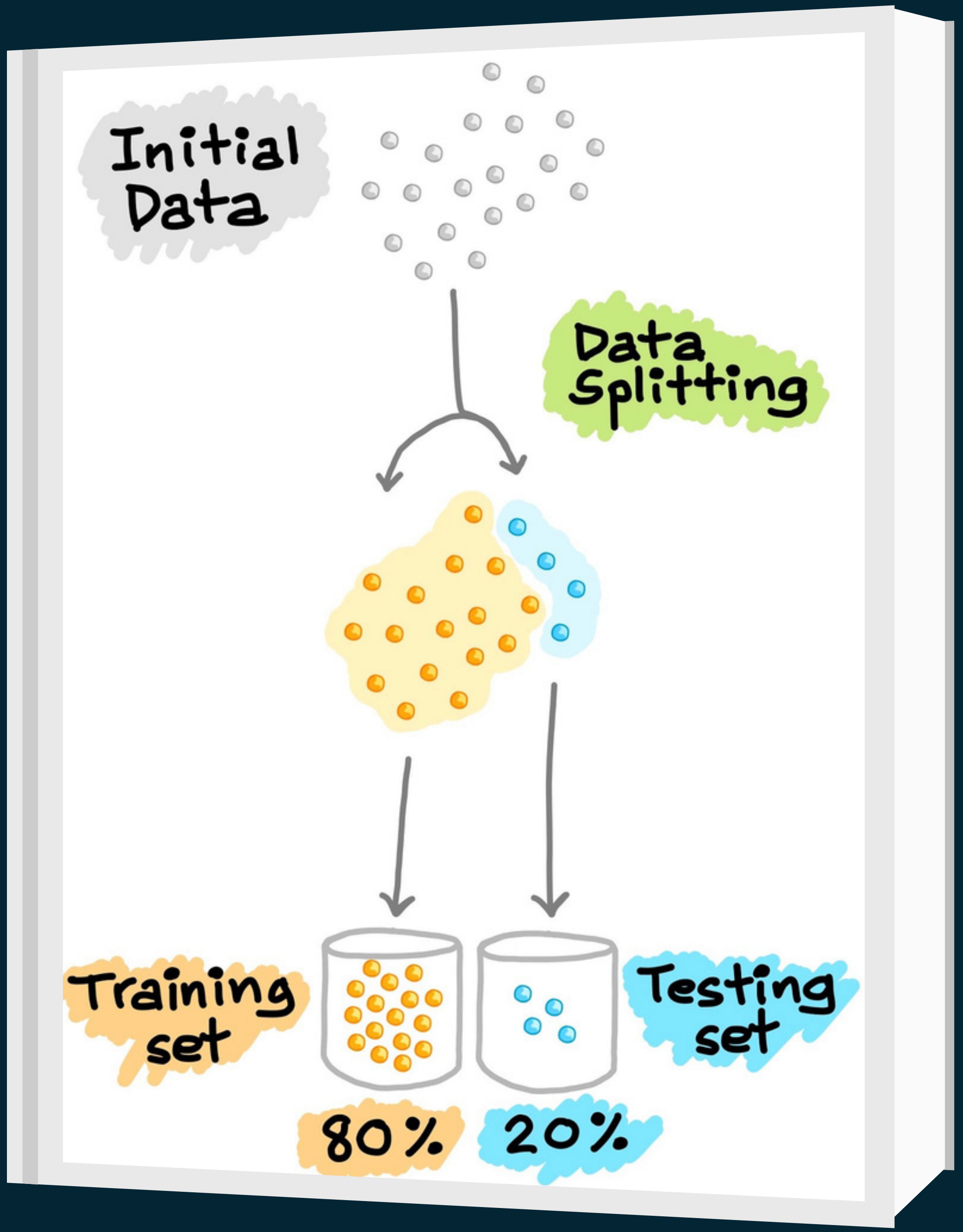
MODEL ARCHITECTURE:

The first step is to design the architecture of the CRNN. This involves selecting the number and types of layers to use, the activation functions, and the optimization algorithm.

The Music Genre Classification model contains five Convolution layers, five MaxPooling layers, one LSTM and two Dense convolution Layers at the end.

LOSS FUNCTION

The loss function is used to measure the difference between the predicted output and the actual output. The goal is to minimize the loss function, which is achieved by adjusting the weights and biases in the network.



OPTIMISATION ALGORITHM:

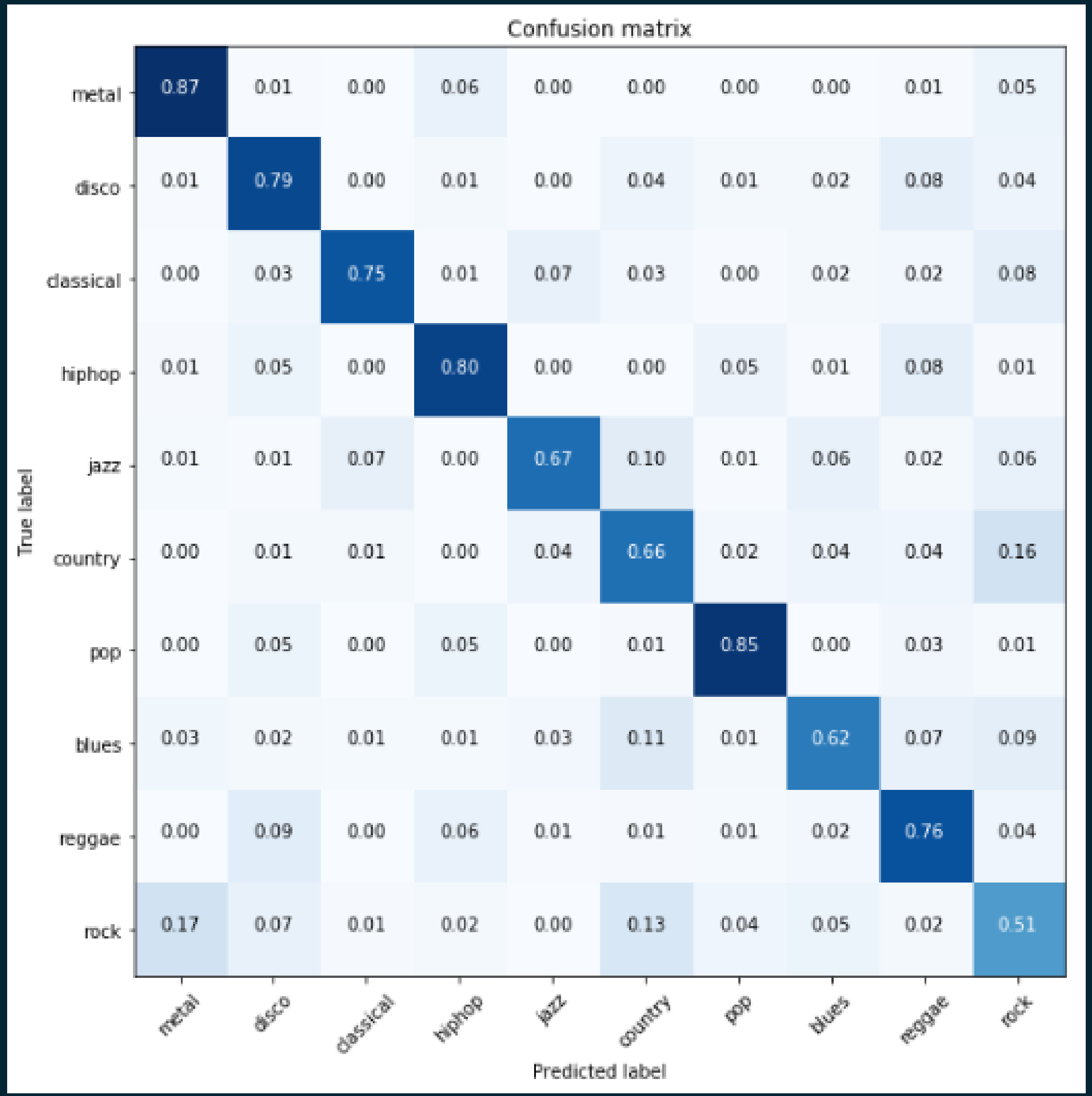
The optimization algorithm is used to update the weights and biases in the network based on the gradients computed from the loss function.

During training, the weights of the convolutional filters and classification model are optimized using backpropagation and gradient descent algorithms to minimize the error between the predicted and actual genre labels.

STAGE 3:

MODEL VALIDATION

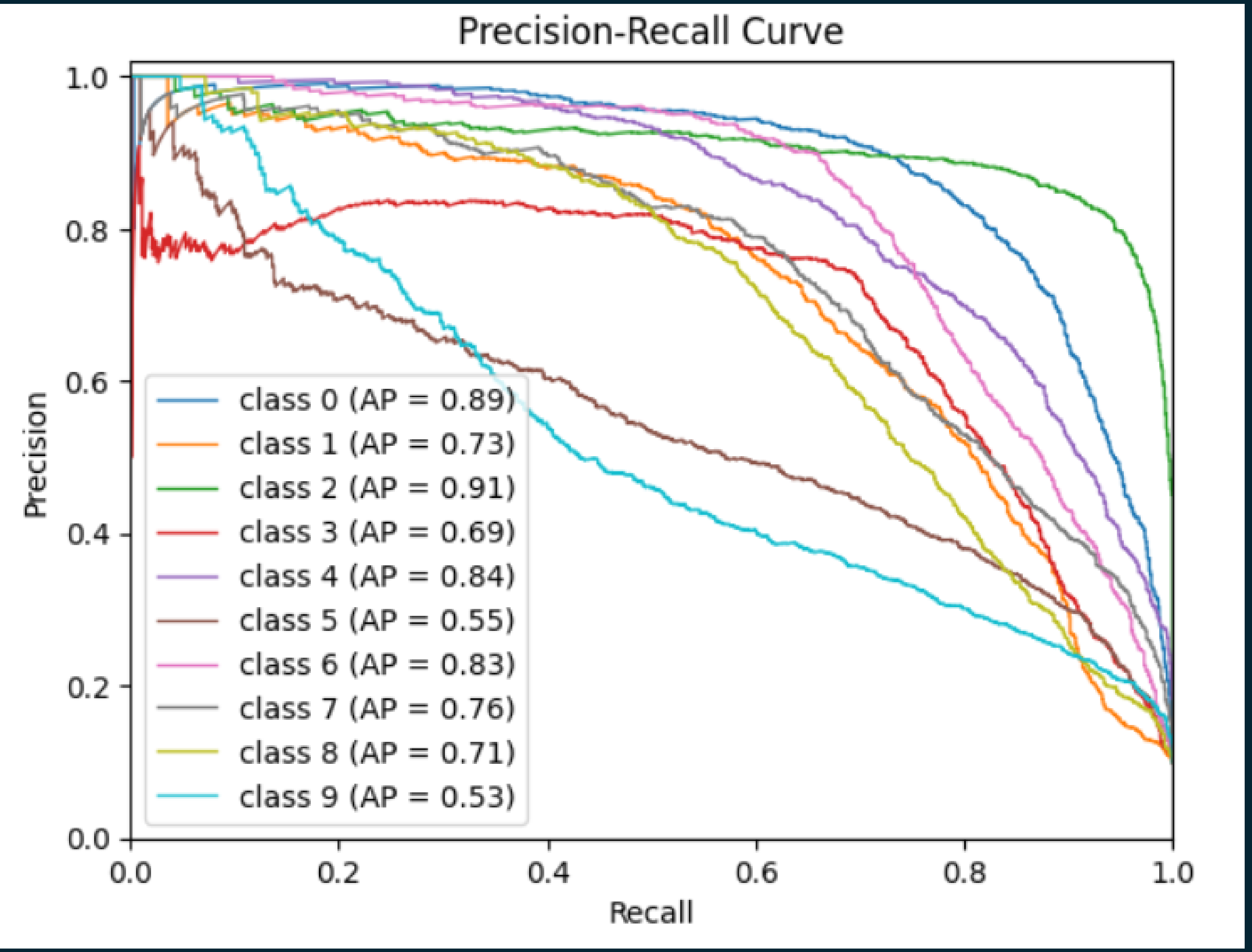
We have used the **Confusion Matrix**, which is a commonly used evaluation metric for multi-class classification tasks, and ensure the accuracy and reliability of the model.



STAGE 3:

MODEL VALIDATION

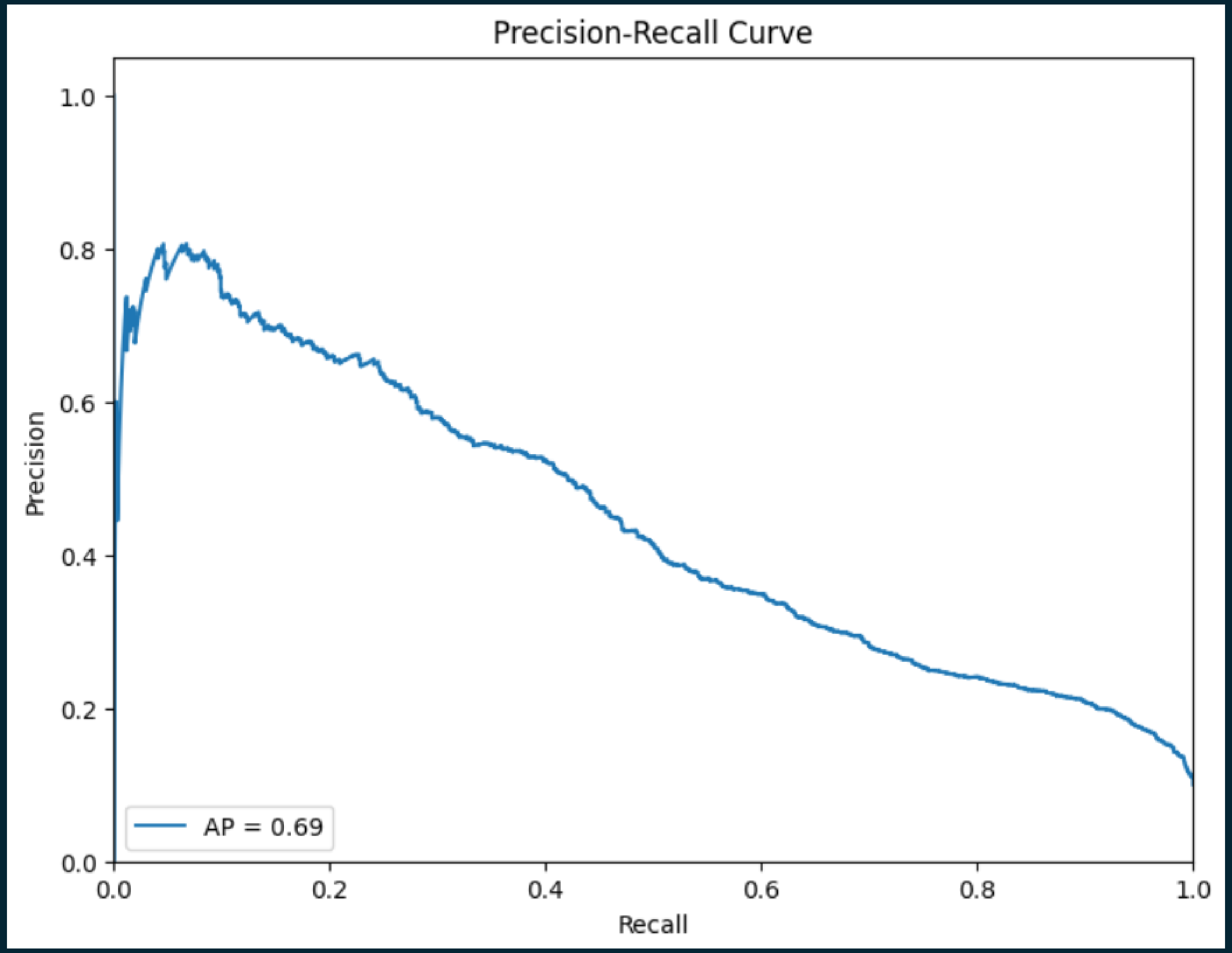
Precision-Recall (PR) curve is a graphical representation of the trade-off between precision and recall for different thresholds of a classifier.



Precision-Recall Curve for individual Categories

STAGE 3:

MODEL VALIDATION



Combined Precision-Realll Curve

STAGE 3:

MODEL VALIDATION

Other Evaluation Metrics:

```
366/366 [==============================] - 3s 7ms/step
MAP@10 Score : 0.7787638595638748
```

```
366/366 [==============================] - 2s 6ms/step
Mean average precision: 0.72
```

```
366/366 [==============================] - 3s 7ms/step
AUC score:  0.9286901567357407
```

```
Accuracy Score (acc) = 0.854
```

STAGE 4:

GENRE PREDICTION

Once a CRNN-based music genre detection system is trained and validated, it can be used to predict the genre(s) of new music tracks with high accuracy.

The prediction process involves extracting features from the raw audio data of the music track, using the same set of convolutional filters that were trained during the model building stage.

The extracted features are then fed into the classification model, which outputs a probability distribution over the different music genres.

Improvemets over the previous Model

Model can now support audio file over 30 seconds

Since these is a limitation of 30 seconds in one prediction, we are dividing the audio file into clips of 30 second and running the model on each one of these audio files. Then we compile the results from all these results to return the final classification.

Improved Accuracy

The recurrent layer of CRNN can help capture long-term dependencies in sequential data, which is very important. For example, in speech recognition, the context of previous words can be important for accurately recognizing the current word.

THANK YOU