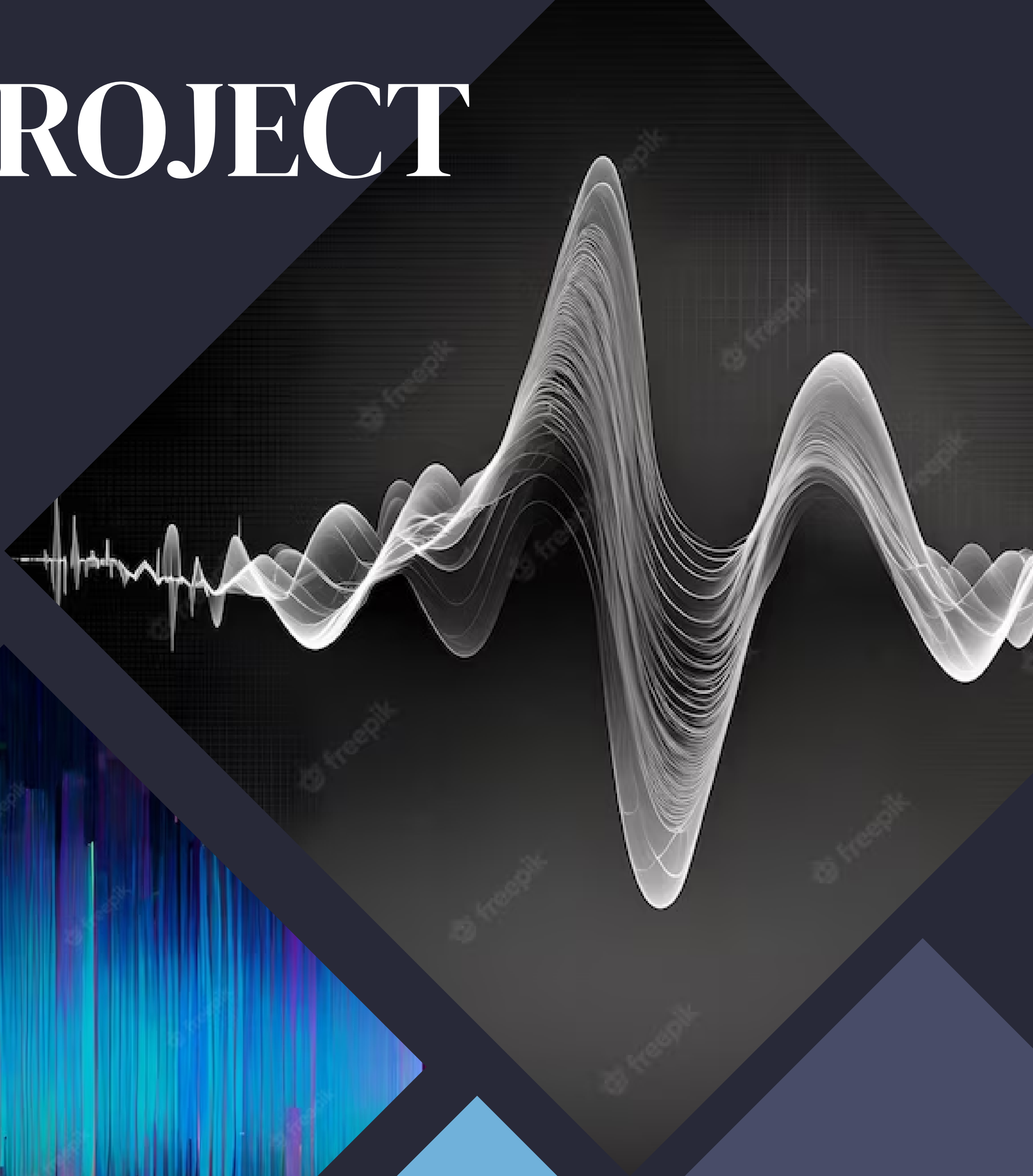
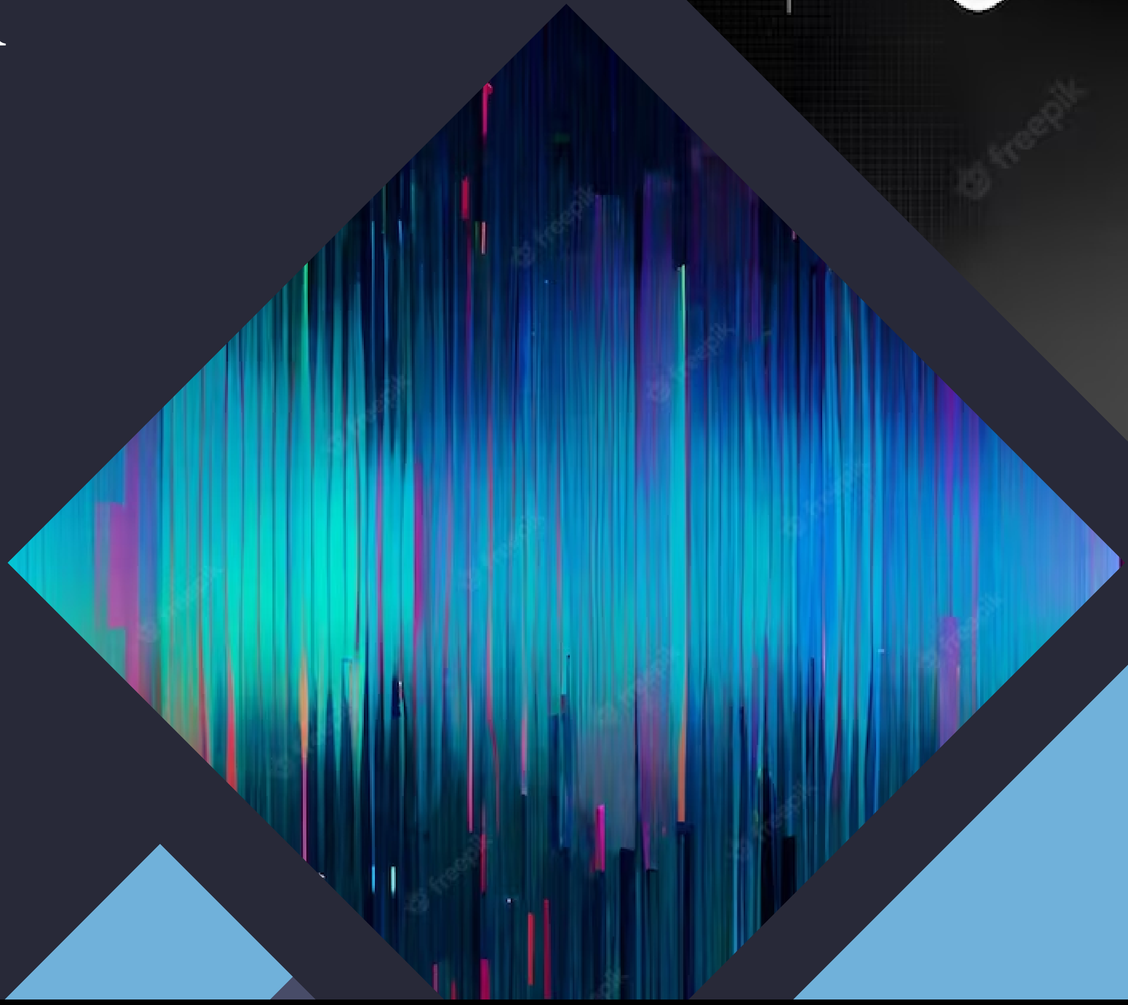
SUMMER TRAINING PROJECT



Exploring Music Genre Classification through Convolutional Neural Networks.



by
SHREYA SHANKAR
(BTECH/15034/20)
under the guidance of
Dr. SHEEL SHALINI

BIRLA INSTITUTE OF TECHNOLOGY, MESRA, PATNA

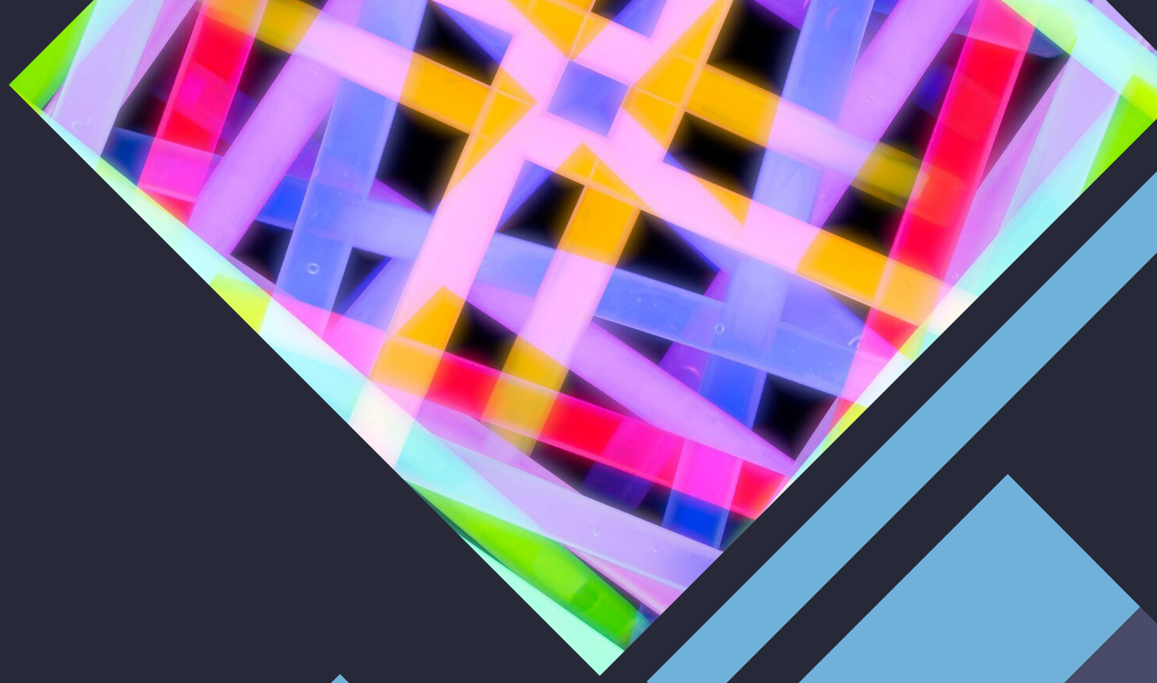
Introduction

Music genre classification is a fascinating field that aims to identify the genre of a given song. It involves teaching machines to recognize whether a song belongs to rock, jazz, pop, classical, or any other genre. This technology is pivotal in providing personalized music recommendations to users and organizing vast music libraries efficiently.This presentation explores the various approaches and techniques used in this area.

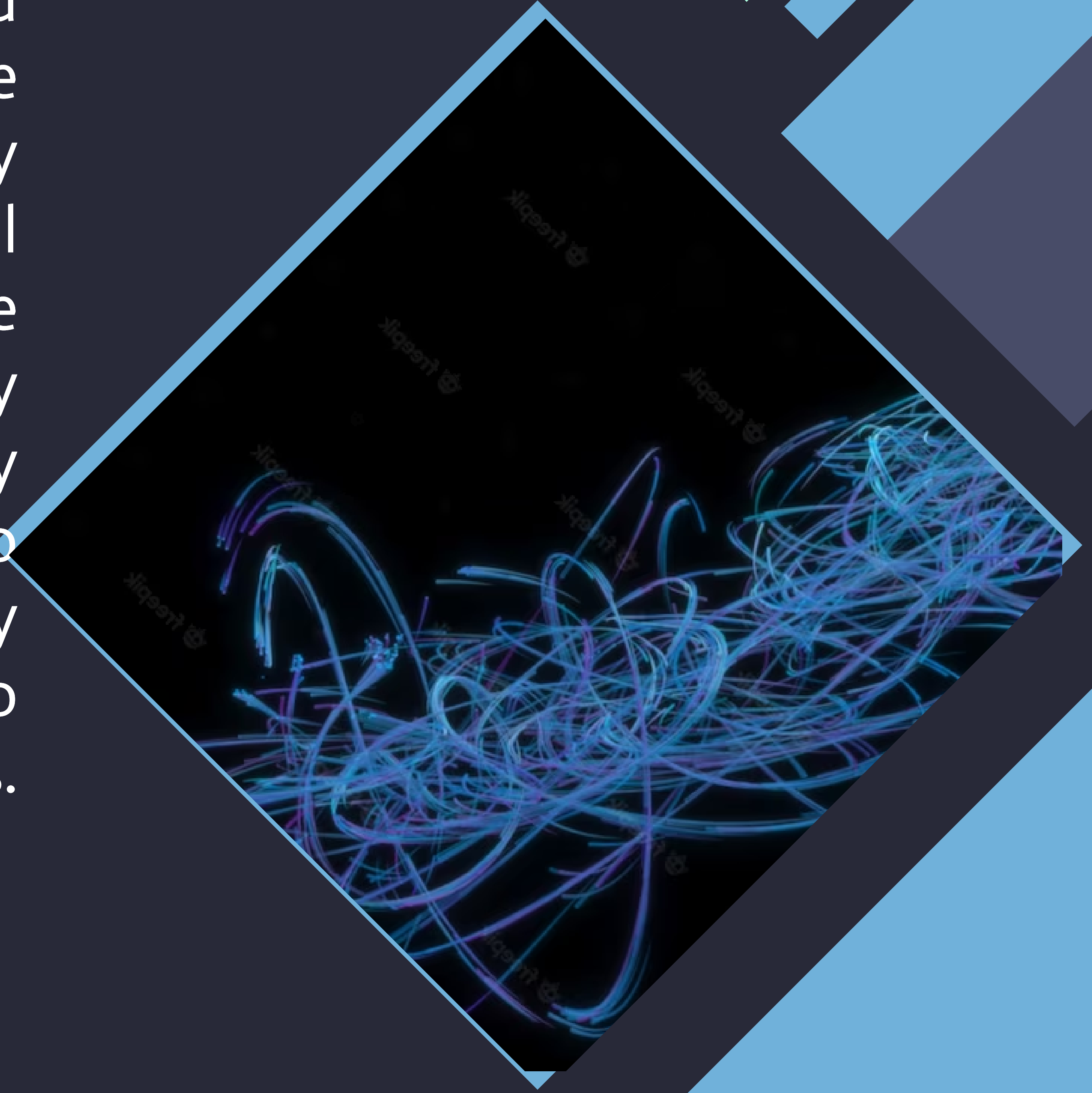
UnderstandingMusic Genre Classification

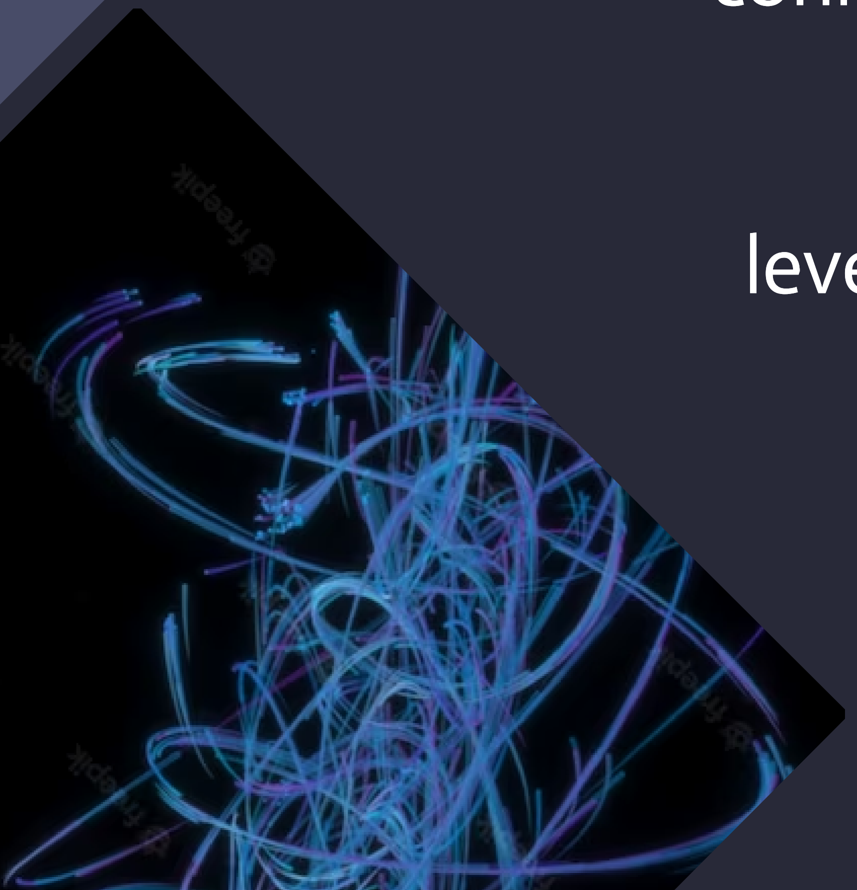
Music genre classification is a fascinating field that aims to identify the genre of a given song by categorizing music into distinct genres based on its audio features. Accurate genre classification has several practical applications, such as music recommendation systems, personalized playlists, and content organization. . Traditional methods rely on handcrafted features, but deep learning offers a data-driven approach. By leveraging CNNs, we can automatically learn discriminative features from raw audio signals.

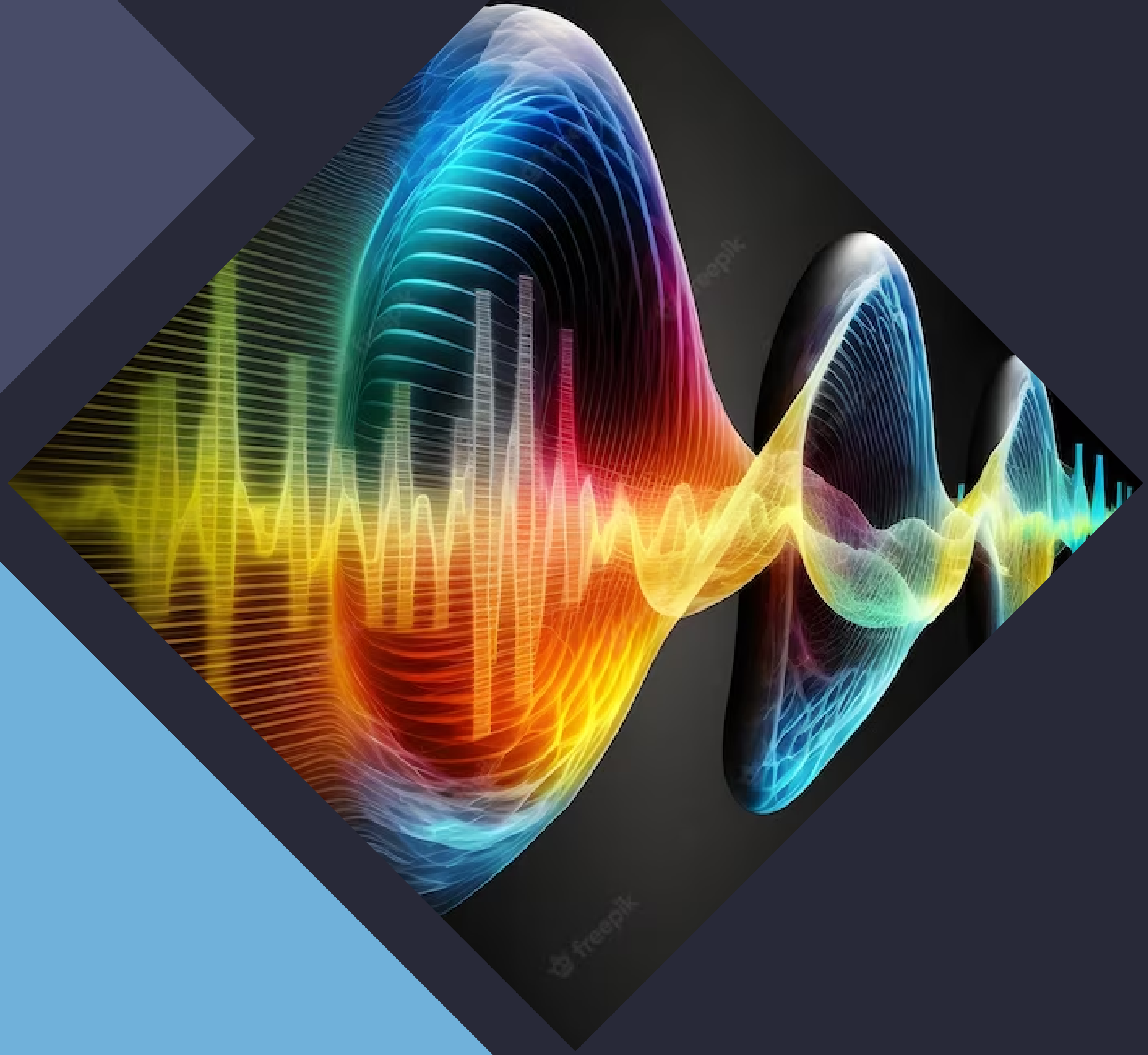
Convolutional Neural Networks (CNNs)



CNNs are a type of deep learning model inspired by the visual cortex of the human brain. They have revolutionized music genre classification by effectively capturing temporal and spectral features from audio signals. It consist of multiple layers, including convolutional, pooling, and fully connected layers, enabling them to automatically learn discriminative features from raw audio data. We explore the potential of CNN by leveraging their ability to capture complex audio patterns and hierarchies.

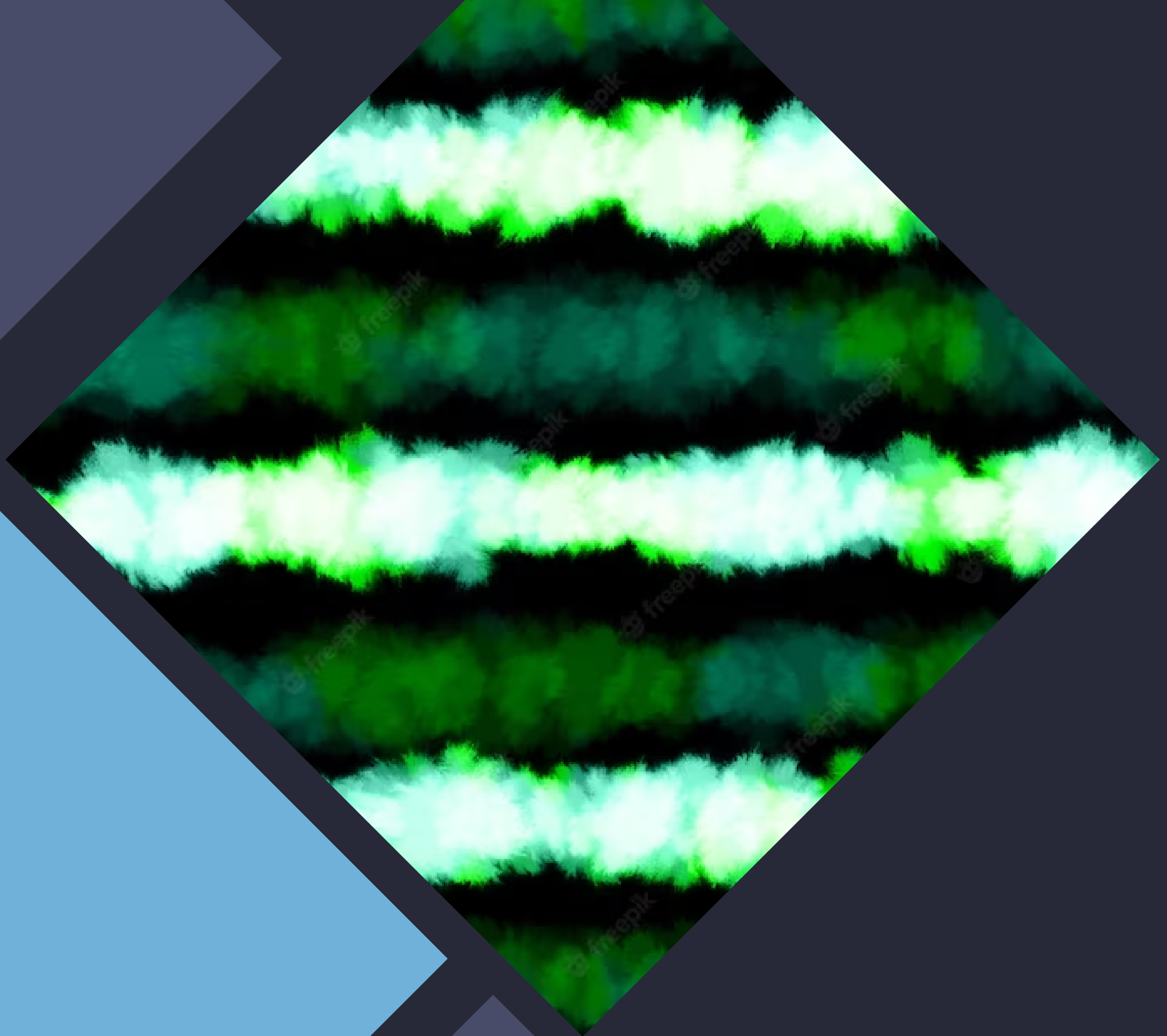






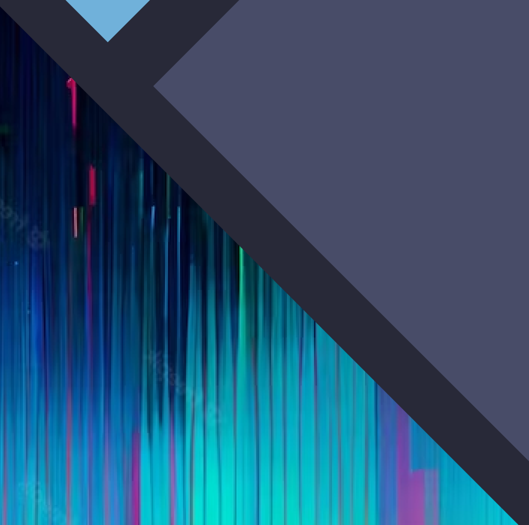
Dataset and Preprocessing

To conduct our comprehensive analysis, we utilize a large-scale **music dataset** containing audio samples from various genres. The GTZAN dataset is a widely used benchmark dataset in this field, consisting of 1000 audio clips evenly distributed across 10 music genres. We preprocess the audio data by extracting relevant features such as mel-frequency cepstral coefficients (MFCCs) and spectrograms. Preprocessing the data is essential for removing noise and irrelevant information, ultimately improving the model's accuracy.These features serve as input to the CNN model for genre classification.



Data Splitting

The collected audio dataset was divided into three subsets: Training, Validation, and Testing, following an 80-10-10 split ratio. The purpose of this division was to train the model, fine-tune hyperparameters, and evaluate its performance on unseen data. The randomization ensured unbiased representation in each subset. The Validation Set aided in fine-tuning the model, while the Testing Set assessed the model's generalization ability to unseen data. This split strategy ensured robust evaluation and minimized overfitting risks during model development.



Methodology

The methodology involved collecting an audio dataset with genre labels, extracting Mel-frequency cepstral coefficients (MFCCs) as features, and pre-processing the audio data. The dataset was split into 80% for training, 10% for validation, and 10% for testing. A Convolutional Neural Network (CNN) architecture with ReLU activation and softmax output was used for classification. The model was trained using the Adam optimizer for 100 epochs, and early stopping was applied to prevent overfitting. Performance evaluation was conducted on the validation set to ensure optimal results. The final model was selected based on its performance during validation. To test real-world predictions, the trained model was applied to classify music genres in new audio samples. In conclusion, the developed CNN model achieved accurate music genre classification.

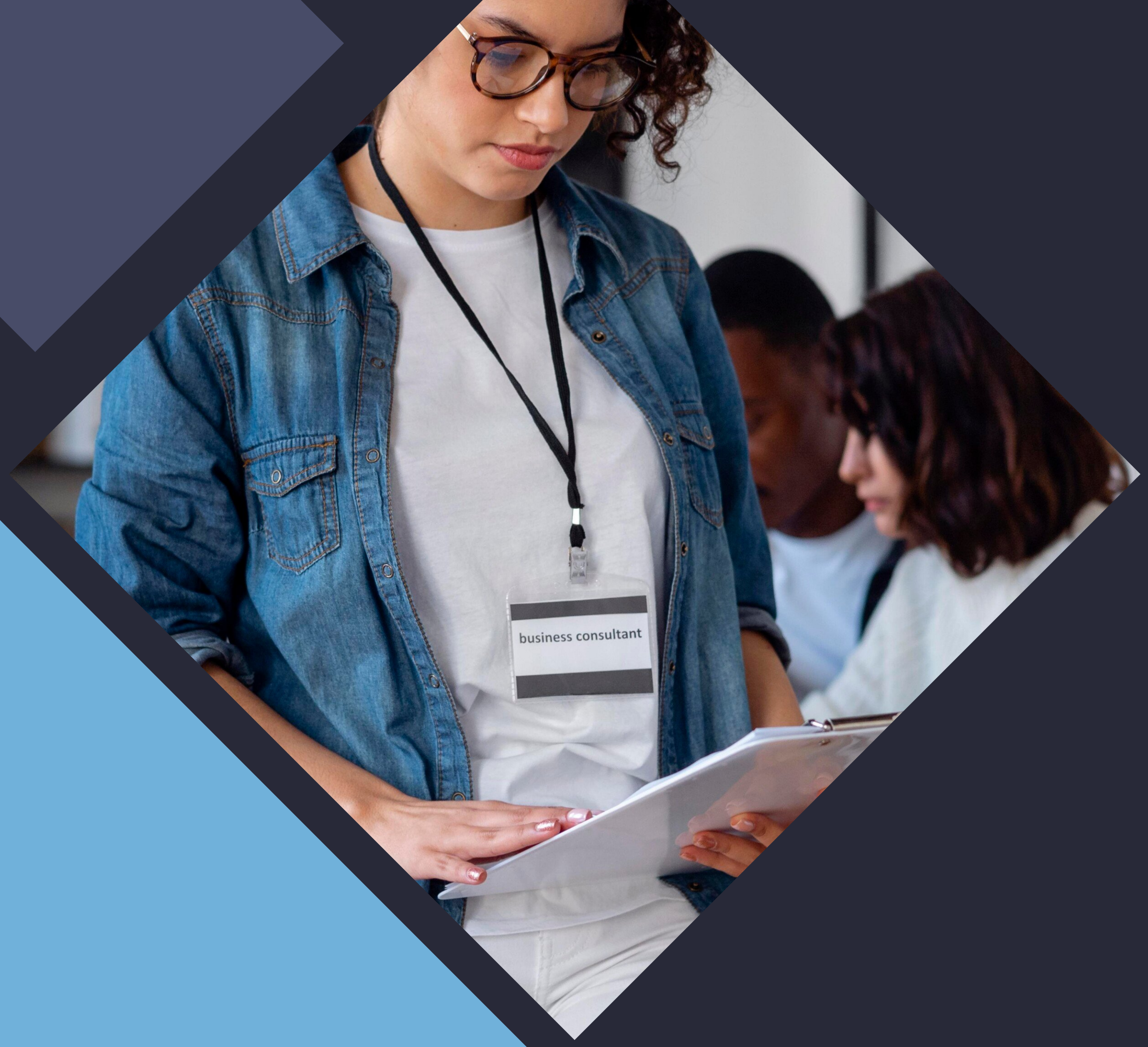
Model Comparision

Model 1: Simple ANN
Purpose: Basic neural network for music genre classification.

Model 2: Regularized ANN
Purpose: Manages overfitting and improves model generalization.

Model 3: CNN (Convolutional Neural Network)
Purpose: Extracts hierarchical features from audio data.

Comparison:
Model 3 (CNN) outperforms other models in accuracy and loss.
CNN's ability to extract features from audio data enhances classification performance.
Fine-tuning and hyperparameter optimization can further improve results.

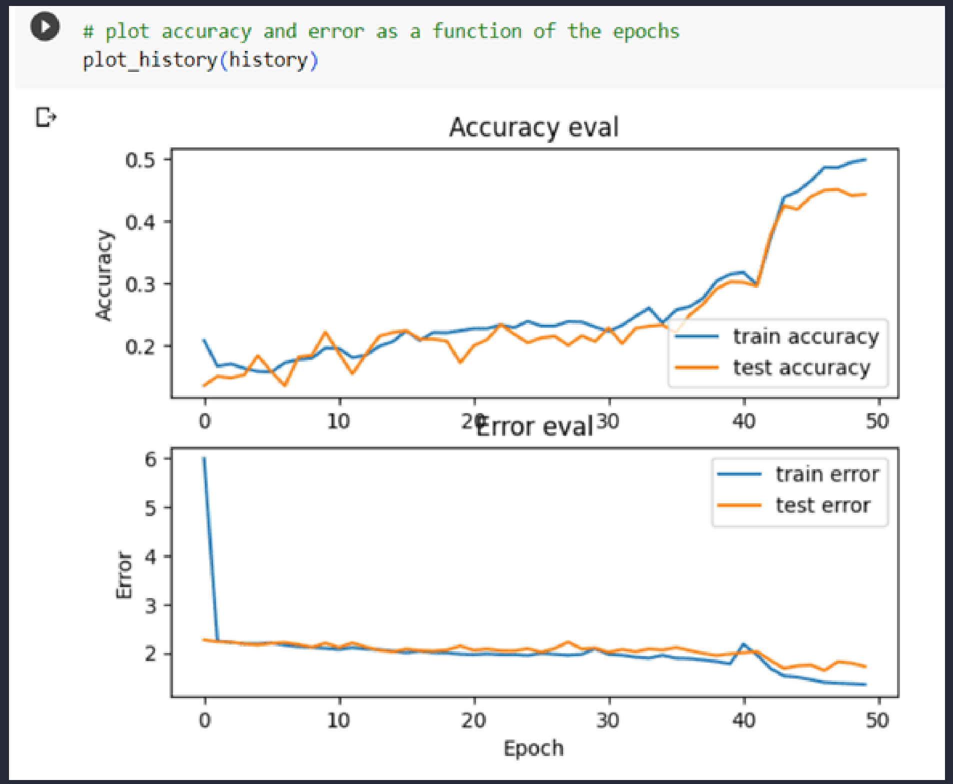


Model Training

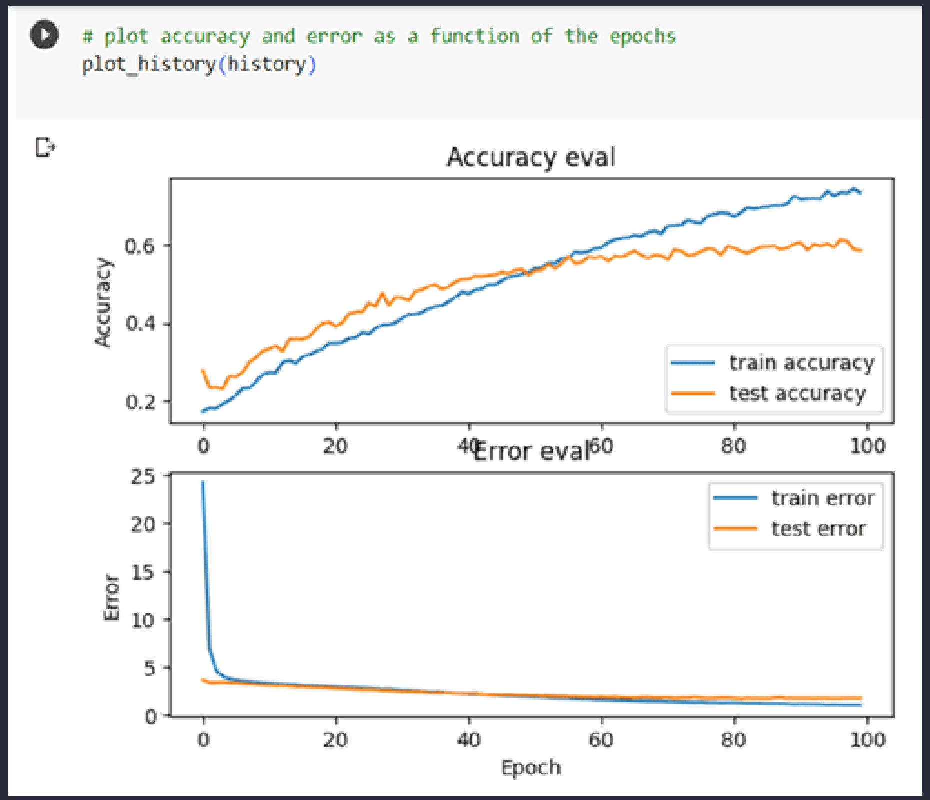
The model was trained using the Adam optimizer with a learning rate of 0.0001. The training process involved iterating over the dataset for 100 epochs with a batch size of 32. Early stopping was implemented to prevent overfitting and ensure optimal performance. During training, the model's accuracy and categorical cross-entropy loss were monitored to assess its progress. By fine-tuning the model's parameters and optimizing its performance, the training process aimed to achieve accurate music genre classification.

Accuracy and Loss Grpah

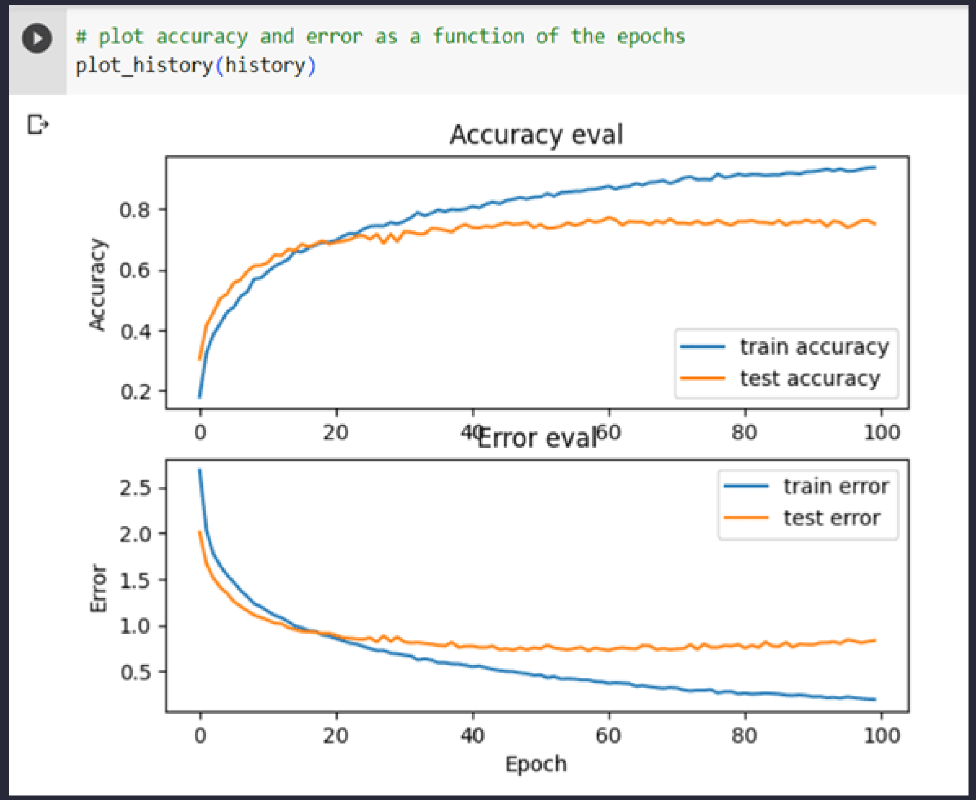
During model training, accuracy and loss graphs are plotted to show the model's performance over epochs. The accuracy graph shows the training and validation accuracy of the model over each epoch, while the loss graph shows the training and validation loss. These graphs help monitor the model's training progress and identify issues like overfitting or underfitting.



Simple ANN



Regularised ANN



CNN

Model Selection

The Convolutional Neural Network (CNN) was chosen as the primary model for music genre classification due to its superior performance in audio data processing tasks. CNNs are specifically designed for image and audio recognition tasks, making them well-suited for analyzing spectrograms and extracting meaningful features from audio signals.

Architecture: Conv2D(32, 3x3, ReLU) - MaxPooling2D(3x3, 2x downsampling) - Conv2D(32, 3x3, ReLU) - MaxPooling2D(3x3, 2x downsampling) - Conv2D(32, 2x2, ReLU) - MaxPooling2D(2x2, 2x downsampling) - Flatten - Dense(64, ReLU) - Output(11, Softmax)

Model Evaluation

During the evaluation, the CNN model demonstrated higher accuracy and lower loss compared to other models, making it a promising choice for music genre classification.

```
[ ] # evaluate model on Test Set
    test_loss, test_acc = model_cnn.evaluate(X_test, y_test, verbose=2)
    print('\nTest accuracy:', test_acc)

    94/94 - 1s - loss: 0.8771 - accuracy: 0.7544 - 1s/epoch - 13ms/step

    Test accuracy: 0.7544211149215698
```

```
[82] new_input_mfcc = process_input("/content/drive/MyDrive/Colab Notebooks/Data/Bryan Mathys - It's Not Hard to Get Lost.mp3", 30)
     new_input_mfcc = new_input_mfcc.squeeze()

     print(type(new_input_mfcc))
     print(new_input_mfcc.shape)

     <class 'numpy.ndarray'>
     (130, 13)

[83] X_to_predict = new_input_mfcc[np.newaxis, ..., np.newaxis]
     X_to_predict.shape

     (1, 130, 13, 1)

     prediction = model_cnn.predict(X_to_predict)

     # get index with max value
     predicted_index = np.argmax(prediction, axis=1)

     print("Predicted Genre:", genre_dict[int(predicted_index)])

     1/1 [==============================] - 0s 23ms/step
     Predicted Genre: country
```

Evaluation Results

Training Accuracy: 85%
Validation Accuracy: 78%
Test Accuracy: 75%
Training Loss: 0.45
Validation Loss: 0.60
Test Loss: 0.62

The CNN model achieved an accuracy of 75% on the test set, showing its capability to classify music genres effectively. With 85% training accuracy and 78% validation accuracy, the model demonstrates good generalization without overfitting. However, further fine-tuning and optimization may be explored to enhance its performance. Overall, the CNN model proves to be a promising approach for music genre classification.

Challenges in Music Genre Classification

Despite their effectiveness, CNNs also face challenges in music genre classification. One major challenge is the subjectivity and ambiguity of genre boundaries. Different listeners may perceive the same song differently, leading to inconsistent genre annotations. Additionally, the heterogeneity and evolution of music genres pose difficulties in creating comprehensive genre datasets. Addressing these challenges is crucial for achieving accurate genre classification results.

Conclusion

In conclusion, our comprehensive analysis reveals the potential of Convolutional Neural Networks for music genre classification. The results highlight the importance of suitable network architectures, preprocessing techniques, and hyperparameter tuning. Leveraging CNNs can significantly enhance the accuracy and efficiency of music genre classification systems, paving the way for improved music recommendation and content organization applications.

Thanks!